ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

Καινοτόμοι Μηχανισμοί Βελτίωσης Απόδοσης και Αποτελεσματικότητας των Βιομετρικών Συστημάτων

Διδακτορική Διατριβή

της

Βασιλικής Η. Ανδρόνικου

Διπλωματούχου Ηλεκτρολόγου Μηχανικού &

Μηχανικού Υπολογιστών Ε.Μ.Π.

Αθήνα, Ιούνιος 2009

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

Καινοτόμοι Μηχανισμοί Βελτίωσης Απόδοσης και Αποτελεσματικότητας των Βιομετρικών Συστημάτων

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Βασιλική Η. Ανδρόνικου

Συμβουλευτική Επιτροπή : Θεοδώρα Α. Βαρβαρίγου
Εμμανουήλ Ν. Πρωτονοτάριος
Γεώργιος Ι. Στασινόπουλος

Εγκρίθηκε από την επταμελή εξεταστική επιτροπή την 29η Ιουνίου 2009.

..... Θεοδώρα Α. Βαρβαρίγου Καθηγήτρια ΕΜΠ	..... Εμμανουήλ Ν. Πρωτονοτάριος Ομότιμος Καθηγητής ΕΜΠ	..... Γεώργιος Ι. Στασινόπουλος Καθηγητής ΕΜΠ
..... Δημήτριος Κουτσούρης Δ.Ε.Π Καθηγητής Ε.Μ.Π	..... Κωνσταντίνα Νικήτα Καθηγήτρια ΕΜΠ	..... Βασίλειος Λούμος Καθηγητής ΕΜΠ
	..... Αναστάσιος Δουλάμης Επίκ. Καθηγητής Πανεπιστημίου Κρήτης	

Αθήνα, Ιούνιος 2009

.....

Βασιλική, Η. Ανδρόνικου

Διδάκτωρ Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

ΠΡΟΛΟΓΟΣ

Η διδακτορική διατριβή που παρουσιάζεται στις επόμενες σελίδες εκπονήθηκε από τον Οκτώβριο του 2004 μέχρι τον Ιούνιο του 2009, στο εργαστήριο Τηλεπικοινωνιών του τομέα Επικοινωνιών, Ηλεκτρονικής και Συστημάτων Πληροφορικής, στη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου.

Κατά την διάρκεια της εκπόνησης αυτής της διατριβής, είχα την ευκαιρία να ασχοληθώ με αρκετά ενδιαφέροντα επιστημονικά θέματα που αφορούν κυρίως στους τομείς της προδιαγραφής, του σχεδιασμού, της υλοποίησης και του ελέγχου βιομετρικών συστημάτων και να αποκτήσω πολύτιμη εμπειρία και γνώσεις.

Θα ήθελα να ευχαριστήσω από τα βάθη της καρδιάς μου την καθηγήτρια μου κα. Θεοδώρα Βαρβαρίγου για την καθοδήγησή της, τις πολύτιμες συμβουλές της και τη σημαντική στήριξη που μου προσέφερε κατά τη διάρκεια αυτής της πορείας μου, καθώς επίσης τους καθηγητές της τριμελούς συμβουλευτικής επιτροπής κ.κ. Εμμανουήλ Πρωτονοτάριο και Γεώργιο Στασινόπουλο.

Επίσης, θα ήθελα να ευχαριστήσω όλους τους συναδέλφους μου στην ερευνητική ομάδα με τους οποίους συνεργάστηκα άψογα και επιτυχώς όλα αυτά τα χρόνια. Ιδιαίτερες ευχαριστίες ωστόσο θα ήθελα να απευθύνω στους στενούς μου συνεργάτες και κυρίως στους Κωνσταντίνο Τσερπέ, Δημήτριο Χάλκο, Ευστάθιο Καραναστάση, Κωστή Χριστογιάννη, Φώτη Αίσωπο, Γαβριήλ Σιδερά, Αναστάσιο και Νικόλαο Δουλάμη με τους οποίους μοιραστήκαμε πάρα πολλές ώρες ερευνητικής εργασίας και αποδοτικής συνεργασίας.

Τέλος, θα ήθελα να ευχαριστήσω τους γονείς μου για την διακριτική καθοδήγησή τους, την θερμή υποστήριξή τους και την πίστη τους σε μένα και στις επιλογές τους καθώς και τους φίλους μου που με στήριξαν όλα αυτά τα χρόνια.

Βασιλική Η. Ανδρόνικου
Ιούνιος 2009

Πίνακας Περιεχομένων

Εικόνες-Σχήματα

Πίνακες

Περίληψη

Η παρατηρούμενη τάση στην αγορά για παροχή εξατομικευμένων υπηρεσιών αλλά και η διαρκώς εντεινόμενη απαίτηση για ενίσχυση της ασφάλειας σε ποικίλους τομείς της ζωής των ανθρώπων τα τελευταία χρόνια οδηγούν σταδιακά στην αναζήτηση νέων, πιο αποτελεσματικών και εύρωστων μηχανισμών, μεθόδων και συστημάτων ταυτοποίησης. Οι βιομετρικές τεχνολογίες βασιζόμενες σε ανθρώπινα χαρακτηριστικά με ιδιαίτερες εγγενείς ιδιότητες, όπως η μοναδικότητα, η παγκοσμιότητα και η συλλεξιμότητα, οι οποίες συνιστούν βασικές απαιτήσεις για την ταυτοποίηση των ανθρώπων, αποτελούν μια συντονισμένη ερευνητική προσπάθεια κάλυψης αυτών των υψηλών αναγκών. Το έντονο εμπορικό αλλά και κυβερνητικό ενδιαφέρον για τις τεχνολογίες αυτές και τις δυνατότητές τους έχει οδηγήσει σε πυρετώδη έρευνα παγκοσμίως όσον αφορά στην αύξηση των επιπέδων ακρίβειας αλλά και τη μείωση του κόστους, ώστε να είναι εφικτή η κάλυψη του διαρκώς αυξανόμενου εύρους εφαρμογών διαφορετικών απαιτήσεων οι οποίες καλούνται να ενσωματώσουν τα συστήματα αυτά για την ασφαλέστερη και αποτελεσματικότερη διαχείριση της ταυτότητας.

Η παρούσα διατριβή αφορά στην παρουσίαση καινοτόμων μηχανισμών τόσο εγγενών στα βιομετρικά συστήματα όσο και σε επίπεδο υποδομών οι οποίοι στοχεύουν στη βελτίωση της αποτελεσματικότητας και της απόδοσης των συστημάτων αυτών. Όσον αφορά στην αποτελεσματικότητα, παρουσιάζεται η μοντελοποίηση των βιομετρικών δεδομένων σε επίπεδο βαθμολογίας σύγκρισης βασισμένης σε Student t mixture models και η εφαρμογή ενός

πιθανοτικού κανόνα απόφασης για τον καθορισμό της εγκυρότητας της ταυτότητας των χρηστών. Ο μηχανισμός αυτός συγκρίνεται με υπάρχουσες μεθόδους συνδυασμού βαθμολογιών σύγκρισης με τα αποτελέσματα προσομοίωσης τα οποία παρατίθενται να αποδεικνύουν την αποτελεσματικότητα και την αποδοτικότητα του προτεινόμενου μηχανισμού. Σημειώνεται ότι ο εν λόγω μηχανισμός είναι ανεξάρτητος από το βιομετρικό σύστημα στο οποίο ενσωματώνεται καθώς και από τους χρησιμοποιούμενους αλγόριθμους εξαγωγής βιομετρικών χαρακτηριστικών.

Στα πλαίσια της έρευνας για την αύξηση της ακρίβειας των αποτελεσμάτων των βιομετρικών συστημάτων, παρουσιάζεται μια καινοτόμος αρχιτεκτονική των βιομετρικών συστημάτων η οποία αποτελεί επέκταση της υπάρχουσας γενικής αρχιτεκτονικής και η οποία στοχεύει στην πιο αποτελεσματική και έξυπνη εκμετάλλευση του πληροφοριακού πλούτου των βιομετρικών δειγμάτων μέσω του συνδυασμού αυτής με αντίστοιχη πληροφορία η οποία αφορά στις ειδικές συνθήκες βέλτιστης απόδοσης των εισηγμένων αλγόριθμων εξαγωγής βιομετρικών χαρακτηριστικών στο βιομετρικό σύστημα. Η μελέτη αυτή επεκτείνεται σε θέματα ασφάλειας και ιδιωτικότητας των βιομετρικών συστημάτων κατά την οποία αναλύονται τα τρωτά σημείων των βιομετρικών συστημάτων και παρατίθεται αντιστοίχηση αυτών με τις κύριες εκφάνσεις της ιδιωτικότητας.

Τέλος, βασιζόμενοι στο γεγονός ότι τα ζητήματα αξιοπιστίας, αποτελεσματικότητας και αποδοτικότητας ενός βιομετρικού συστήματος είναι άρρηκτα συνδεδεμένα με τη διαθεσιμότητα των βιομετρικών δεδομένων και την ταχύτητα πρόσβασης σε αυτά, παρουσιάζονται καινοτόμοι μηχανισμοί δημιουργίας και διατήρησης αντιγράφων σε κατανεμημένο περιβάλλον, η αξιολόγηση των οποίων γίνεται με έμφαση στην απαιτούμενη υπολογιστική ισχύ αυτών και την ταχύτητα εκτέλεσής τους χωρίς μείωση της αποτελεσματικότητά τους σε μη αποδεκτά επίπεδα.

Abstract

The current market trend for the provision of personalised services as well as the intensifying need for reinforcing security in various sectors over the past years are gradually leading towards the quest for novel, more effective and robust identification mechanisms, methods and systems. Biometric technologies, being based on human characteristics with special innate properties, such as uniqueness, universality and collectiveness, which comprise important requirements for person identification, are a joint multidisciplinary research effort for covering these high needs. The intense commercial and governmental interest for these technologies and their capabilities has led to worldwide feverish research regarding boosting their accuracy while reducing their cost, so that they are able to cover the continuously increasing range of applications of varying requirements aiming at providing more secure and effective identity management.

This thesis presents innovative mechanisms which are either inextricable parts of biometric systems or focus on the infrastructural requirements in order to improve the effectiveness and the efficiency of these systems. As far as effectiveness is concerned, we present an innovative modeling of biometric data at score-level based on Student t mixture models and the application of a likelihood-based decision rule. This mechanism is compared against existing methods for score-level fusion with the simulation results proving the effectiveness and the efficacy of the proposed method. It should be noted that this mechanism is independent of the biometric system in which it is incorporated as well as the biometric feature extraction algorithms used.

In the context of increasing the accuracy of the biometric systems, a novel biometric systems architecture is presented which is an extension of the existing generic architecture of biometric systems and which focuses on the more effective and intelligent exploitation of the informational wealth of biometric data through its combination with the respective information regarding the specific conditions under which the incorporated biometric feature extraction algorithms perform best. This study expands towards security and privacy aspects of biometric systems during which we analyse the main points of attack within a biometric systems and map them with the major privacy aspects.

Finally, based on the fact that reliability, effectiveness and efficiency issues of a biometric system are tightly linked with the availability of the biometric data and the access latency to the latter, we present novel replica creation and maintainance mechanisms within a distributed environment, with their evaluation being based on the required computational power and their execution time while their effectiveness lies within acceptable levels.

1 *Εισαγωγή*

Στο κεφάλαιο αυτό παρουσιάζεται η έννοια των βιομετρικών συστημάτων και στη συνέχεια αναλύεται η διάρθρωση της παρούσης Διατριβής και περιγράφεται συνοπτικά το περιεχόμενο των κεφαλαίων.

1.1 *Έννοια Βιομετρικών Συστημάτων*

Οι βιομετρικές τεχνολογίες έχουν προκαλέσει πλήθος συζητήσεων και διαφωνιών τα τελευταία χρόνια καθώς τόσο οι κυβερνήσεις όσο και ο επιχειρηματικός κόσμος σταδιακά δείχνουν αυξανόμενο ενδιαφέρον στην υιοθέτηση τεχνολογιών βασισμένων σε βιομετρικά χαρακτηριστικά σε μεγάλης κλίμακας εφαρμογές σχετικές με την ταυτοποίηση και πιστοποίηση της ταυτότητας των πολιτών, όπως τα βιομετρικά διαβατήρια. Το κυβερνητικό και επιχειρηματικό αυτό ενδιαφέρον έχει ενδυναμωθεί από το αυξανόμενο εμπορικό ενδιαφέρον για τις βιομετρικές εφαρμογές και τις δυνατότητες αυτών καθώς και από την πυρετώδη ερευνητική δραστηριότητα στον εν λόγω χώρο η οποία τα τελευταία χρόνια έχει προσφέρει σημαντική βελτίωση στις επιδόσεις των συστημάτων. Η χρήση των βιομετρικών χαρακτηριστικών για ταυτοποίηση και πιστοποίηση της ταυτότητας ενός ατόμου δεν αποτελεί όμως κάποια

επαναστατική μέθοδο των τελευταίων ετών, αλλά βασίζεται στην ίδια τη λειτουργία του ανθρώπου να αναγνωρίζει άτομα από το πρόσωπο, τη φωνή, το περπάτημα.

Ο όρος *βιομετρικός* (biometric : bio + metric) προέρχεται από τον συνδυασμό των ελληνικών λέξεων «*βίος*» που σημαίνει ζωή και «*μετρικός*» που σημαίνει αυτός που μετράει. Έτσι, οι βιομετρικές τεχνολογίες βασίζονται σε ανθρώπινα χαρακτηριστικά με συγκεκριμένες ιδιότητες όπως θα δούμε και στο Κεφάλαιο 3 της παρούσης Διατριβής τα οποία επιτρέπουν με κατάλληλη λήψη, αποτύπωση, ανάλυση και επεξεργασία την ταυτοποίηση του ατόμου στο οποίο ανήκουν.

Σύμφωνα με τους Jain et al [1] ένα βιομετρικό σύστημα αποτελεί επί της ουσίας ένα σύστημα ταυτοποίηση προτύπων το οποίο λειτουργεί με την απόκτηση βιομετρικών δεδομένων από ένα άτομο, την εξαγωγή ενός συνόλου χαρακτηριστικών γνωρισμάτων από τα δεδομένα αυτά, και τη σύγκριση αυτού του συνόλου με ένα προαποθηκευμένο σύνολο πρότυπων χαρακτηριστικών γνωρισμάτων.

Ο Bromme [2] θέλοντας να αποτυπώσει την έννοια των *βιομετρικών συστημάτων στον χώρο της ασφάλειας των πληροφοριακών συστημάτων* αναφέρει ότι αποτελούν μια μέθοδο ταυτοποίηση του ατόμου βασισμένη στην απόκτηση ανθρώπινων βιολογικών χαρακτηριστικών μέσω αισθητήρων, στη μέτρηση των βιομετρικών χαρακτηριστικών (ανεπεξέργαστα δεδομένα και δεδομένα βαθμονόμησης αισθητήρα), στον υπολογισμό των βιομετρικών δεδομένων και των βιομετρικών προτύπων, και στην πιστοποίηση και ταυτοποίηση με βάση βιομετρικά πρότυπα σε σχέση με μαθηματικούς ορισμούς των μετρικών και των μετρικών χώρων.

Το Biometric Consortium [3] τα ορίζει ως αυτοματοποιημένες μεθόδους ταυτοποίησης ενός ατόμου με βάση ένα φυσιολογικό και συμπεριφορικό χαρακτηριστικό. Μεταξύ αυτών των μετρηθέντων χαρακτηριστικών είναι το πρόσωπο, τα δακτυλικά αποτυπώματα, η γεωμετρία του χεριού, η γραφή, η ίρις και η φωνή.

1.2 *Όραμα και Ανοιχτά Τεχνολογικά Θέματα*

Τα τελευταία χρόνια η έρευνα στο πεδίο των βιομετρικών συστημάτων προχωρά με γοργούς ρυθμούς τόσο σε Ευρωπαϊκό όσο και σε παγκόσμιο επίπεδο με πλήθος ερευνητών από διαφορετικούς χώρους, όπως τεχνολογικό, νομικό, κοινωνικό, οικονομικό και κυβερνητικό. Στα πλαίσια αυτής της πυρετώδους ερευνητικής δραστηριότητας και του διαρκώς διευρυνόμενου πεδίου εφαρμογής των βιομετρικών συστημάτων, ομάδες έμπειρων επιστημόνων και ερευνητών έχουν διατυπώσει και εργάζονται αναφορικά με το όραμα και τα ανοιχτά τεχνολογικά θέματα των βιομετρικών τεχνολογιών.

Δεδομένων των διαφορετικών τεχνολογιών που βρίσκονται υπό την ομπρέλα των βιομετρικών συστημάτων, όπως είναι η ταυτοποίηση προσώπου, φωνής, περπατήματος, τρόπου υπογραφής, οι ειδικοί σε κάθε χώρο εργάζονται εντατικά προκειμένου να αυξήσουν την αποτελεσματικότητα των αλγορίθμων ή να αναπτύξουν νέους πιο αποδοτικούς τόσο όσον αφορά στις επεξεργαστικές απαιτήσεις όσο και στην ακρίβεια των αποτελεσμάτων αυτών. Παράλληλα, δεδομένου του πλούτου των βιομετρικών χαρακτηριστικών του ανθρώπου έρευνα πραγματοποιείται προκειμένου να προσδιοριστούν νέα βιομετρικά χαρακτηριστικά επιζητώντας τόσο την ενίσχυση της ακρίβειας των βιομετρικών συστημάτων όσο και τη μείωση του κόστους αυτών καθώς και της επεμβατικότητάς τους στην ανθρώπινη ζωή. Έτσι, για παράδειγμα, ένα νέο βιομετρικό χαρακτηριστικό το οποίο είναι υπό έρευνα αυτήν την περίοδο είναι οι ωτοακουστικές εκπομπές (Otoacoustic Emissions - OAEs), οι οποίες βασίζονται στους ήχους που παράγει το αυτί ανταποκρινόμενο σε ακουστικό ερέθισμα. Σύμφωνα με το Βρετανικό Ερευνητικό Συμβούλιο Μηχανικής και Φυσικών Επιστημών (Engineering and Physical Sciences Research Council (EPSRC)) [4] η πηγή αυτών των εκπομπών έγκειται στη λειτουργία του ανθρώπινου σώματος να ενισχύει ήχους χαμηλού επιπέδου και σύμφωνα με τις μελέτες τους οι εκπομπές αυτές είναι

μοναδικές ανάμεσα στους ανθρώπους. Οι εκπομπές αυτές είναι ήχοι παραγόμενοι στο ανθρώπινο αυτί οι οποίοι πηγάζουν από τον σπειροειδή κοχλία στο εσωτερικό του αυτιού.

Παράλληλα και στα πλαίσια της υιοθέτησης των βιομετρικών συστημάτων για τη βελτίωση συστημάτων ασφαλείας, σημαντικές προσπάθειες λαμβάνουν χώρα για την ανάπτυξη τέτοιων συστημάτων πραγματικού χρόνου. Δεδομένης της απαιτούμενης υπολογιστικής ισχύος αλλά και των θεμάτων ιδιωτικότητας και ασφάλειας των δεδομένων που είναι άρρηκτα συνδεδεμένα με τη χρήση των βιομετρικών συστημάτων σε εφαρμογές μεγάλης κλίμακας, το ερευνητικό ενδιαφέρον έχει στραφεί προς την εκμετάλλευση υπαρχουσών υποδομών καθώς και την περαιτέρω ανάπτυξη αυτών προκειμένου να είναι σε θέση να υποστηρίξουν τις αυξημένες λειτουργικές και μη απαιτήσεις των βιομετρικών εφαρμογών.

1.3 *Συνεισφορά – Καινοτομία Διατριβής*

Η Διατριβή επικεντρώνεται στη μελέτη των βιομετρικών συστημάτων (biometric systems) και συγκεκριμένα στη βελτίωση της απόδοσης και της αποτελεσματικότητας αναφορικά με την πρόταση καινοτόμων μηχανισμών συνδυασμού των αποτελεσμάτων διαφορετικών αλγορίθμων όσο και με την ανάπτυξη αποδοτικών τεχνικών διαχείρισης βιομετρικών δεδομένων σε κατανεμημένο περιβάλλον. Σε ένα τέτοιο πλαίσιο σχεδιάζεται και προτείνεται μια καινοτόμος αρχιτεκτονική βιομετρικών συστημάτων η οποία στηρίζεται σε μεταπεριγραφείς τόσο των εφαρμοζόμενων αλγόριθμων εξαγωγής βιομετρικών χαρακτηριστικών όσο και των προς ανάλυση βιομετρικών δεδομένων. Επιπροσθέτως, στα πλαίσια των πολυτροπικών και διατροπικών βιομετρικών συστημάτων παρουσιάζεται ένας αποτελεσματικός μηχανισμός συνδυασμού των βαθμολογιών των εφαρμοζόμενων αλγόριθμων σύγκρισης βιομετρικών δειγμάτων και βιομετρικών προτύπων αναφοράς.

Ο μηχανισμός αυτός βασίζεται στην εφαρμογή των Student t mixture models με τα οποία μοντελοποιούνται οι κατανομές των έγκυρων και μη χρηστών ενός βιομετρικού συστήματος με την αρχικοποίηση των παραμέτρων των μοντέλων να λαμβάνει χώρα με την εφαρμογή του αλγόριθμου EM (Expectation Maximisation). Με βάση τη μοντελοποίηση αυτή πραγματοποιείται συνδυασμός σε επίπεδο βαθμολογίας (score-level fusion) με τον κανόνα απόφασης να βασίζεται στο θέωρημα Neyman-Pearson για την δοκιμή μιας υπόθεσης έναντι της αντίθετης αυτής. Ο μηχανισμός αυτός προσομοιώνεται και εξετάζεται πειραματικά για διαφορετικά δεδομένα και με διαφορετικές παραμέτρους των μοντέλων και τα αποτελέσματα αυτού συγκρίνονται με γνωστούς αντίστοιχους αλγόριθμους. Παράλληλα, εξετάστηκαν θέματα που αφορούν σε ζητήματα ασφάλειας και ιδιωτικότητας όσον αφορά στη λειτουργία των βιομετρικών συστημάτων και στις διάφορες εφαρμογές αυτών. Τέλος, στα πλαίσια της μελέτης της αποδοτικότητας των βιομετρικών συστημάτων αναπτύχθηκε ένα υποσύστημα διαχείρισης αντιγράφων δεδομένων σε κατανεμημένο περιβάλλον με βάση τις ιδιαίτερες απαιτήσεις των βιομετρικών συστημάτων. Οι παρουσιαζόμενοι αλγόριθμοι είναι παραλληλίσιμοι και επιτρέπουν με τον τρόπο αυτόν την αποδοτικότερη εκτέλεση τους με την εισαγωγή τους σε ένα περιβάλλον πλέγματος, ενώ παρουσιάζονται εναλλακτικές καινοτόμες υβριδικές μέθοδοι για την επίλυση των επιμέρους προβλημάτων διαχείρισης αντιγράφων με στόχο την ταχύτερη εκτέλεση του συστήματος.

1.4 *Οργάνωση της Διατριβής*

Η Διατριβή αποτελείται από 8 κεφάλαια. Το πρώτο κεφάλαιο παρουσιάζει την έννοια του βιομετρικού συστήματος καθώς και τις σύγχρονες τάσεις - το όραμα –στον χώρο αυτόν.

Επιπλέον, αναφέρονται τα ανοιχτά τεχνολογικά θέματα και βάσει αυτών παρουσιάζεται η συνεισφορά και η καινοτομία της διατριβής.

Στο δεύτερο κεφάλαιο γίνεται μια εισαγωγή στην έννοια της διαχείρισης ταυτότητας καθώς και τα ζητήματα ταυτότητας και ταυτοποίησης ενώ παρατίθενται οι κυριότεροι κίνδυνοι οι οποίοι ελλοχεύουν κατά τη διαδικασία της ταυτοποίησης και αναλύεται η έννοια της ιδιωτικότητας αναφορικά με την ταυτοποίηση.

Στο τρίτο κεφάλαιο δίνεται μια αναλυτική περιγραφή των βιομετρικών συστημάτων, η οποία περιλαμβάνει τις βασικές κατηγοριοποιήσεις αυτών καθώς και αναλυτική παρουσίαση παραδειγμάτων βιομετρικών συστημάτων, ενώ παρουσιάζονται σύγχρονες βιομετρικές εφαρμογές.

Στο τέταρτο κεφάλαιο συζητούνται στοιχεία αρχιτεκτονικής των βιομετρικών συστημάτων. Γίνεται αναφορά στις φάσεις ενός βιομετρικού συστήματος καθώς και στα επιμέρους υποσυστήματα ενός βιομετρικού συστήματος και αναλύονται τα κυριότερα σφάλματα των βιομετρικών συστημάτων καθώς και οι κυριότεροι δείκτες αξιολόγησης αυτών.

Στο πέμπτο κεφάλαιο αναλύονται οι έννοιες της ασφάλειας και της ιδιωτικότητας και παρουσιάζεται η μελέτη μας σχετικά με τα κύρια τρωτά σημεία των βιομετρικών δεδομένων και οι παραβιάσεις στην ιδιωτικότητα των ατόμων ακολουθούμενη από μια καινοτόμο αντιστοίχηση αυτών. Επιπροσθέτως, περιγράφεται η δυνατότητα των βιομετρικών συστημάτων να αποτελέσουν τεχνολογίες διασφάλισης της ιδιωτικότητας.

Στο έκτο κεφάλαιο περιγράφεται η προτεινόμενη καινοτόμος αρχιτεκτονική βιομετρικών συστημάτων με στόχο τη βελτίωση της ακρίβειας των αποτελεσμάτων μέσω της εκμετάλλευσης μεγαλύτερου μέρους του πληροφοριακού πλούτου στο υπό ανάλυση ληφθέν βιομετρικό δείγμα και κατάλληλου έξυπνου συνδυασμού αυτού και παρουσιάζονται καινοτόμοι μηχανισμοί

συνδυασμού των αποτελεσμάτων σύγκρισης των υπό ανάλυση βιομετρικών δεδομένων με τα βιομετρικά δεδομένα αναφοράς. Τέλος, παρατίθεται η αξιολόγηση αυτών των μηχανισμών.

Στο έβδομο κεφάλαιο περιγράφονται προτεινόμενοι καινοτόμοι μηχανισμοί δημιουργίας και διατήρησης αντιγράφων βιομετρικών δεδομένων σε κατανεμημένο περιβάλλον καθώς και η αξιολόγηση αυτών.

Τέλος, στο όγδοο κεφάλαιο περιλαμβάνεται η σύνοψη της διατριβής και τα συμπεράσματα που εξήχθησαν κατά την εκπόνησή της καθώς και η κύρια συνεισφορά της στο χώρο των βιομετρικών συστημάτων και συζητούνται θέματα μελλοντικής εργασίας και επέκτασης των ερευνητικών αποτελεσμάτων.

2 *Διαχείριση Ταυτότητας*

Στο κεφάλαιο αυτό παρουσιάζεται η έννοια της διαχείρισης ταυτότητας και οι διάφορες διαστάσεις αυτής. Αναλύονται οι έννοιες της ταυτότητας και της ταυτοποίησης, ενώ περιγράφεται το εύρος των εφαρμογών τους. Τέλος, παρατίθενται οι κυριότεροι κίνδυνοι οι οποίοι είναι συνδεδεμένοι με την ταυτοποίηση και την πιστοποίηση των ατόμων και παρουσιάζεται αναλυτικά η έννοια της ιδιωτικότητας καθώς και έννοιες και μηχανισμοί οι οποίοι στοχεύουν στην προστασία της.

2.1 *Η έννοια της διαχείρισης ταυτότητας*

Η *διαχείριση ταυτότητας* αφορά στη διαχείριση του κύκλου ζωής της ταυτότητας ενός υποκειμένου ή ενός αντικειμένου. Ο όρος αυτός εμπεριέχει πολλά διαφορετικά ζητήματα σχετικά με την ταυτότητα, όπως είναι η εξακρίβωσή της (διαδικασία ταυτοποίησης χρήστη) αλλά και θέματα που έχουν να κάνουν με τη διαφύλαξη, την εμπιστευτικότητα, την ασφάλεια, την πρόσβαση, τη διαλειτουργικότητα. Ωστόσο, η διαχείριση της ταυτότητας είναι ένα ζήτημα πιο ευρύ και οριζόντιο που ανακύπτει συνολικότερα και έχει να κάνει με το σύνολο των

πληροφοριών που καθορίζουν μοναδικά μια οντότητα σε ένα πολλαπλό και ευρύ φάσμα δραστηριοτήτων με τρόπο που να είναι άμεσα αναγνωρίσιμη.

Στον επιχειρηματικό και τεχνολογικό κόσμο ο όρος αυτός αφορά σε ένα σύνολο διαδικασιών, μεθόδων και τεχνολογιών που διευκολύνουν και ελέγχουν την πρόσβαση χρηστών σε εφαρμογές και πόρους με παράλληλη προστασία εμπιστευτικής προσωπικής και επιχειρησιακής πληροφορίας από μη εξουσιοδοτημένους χρήστες.

Στον επιχειρηματικό κόσμο συγκεκριμένα η διαχείριση ταυτότητας συνιστά σύγκλιση τεχνολογικών και επιχειρησιακών διαδικασιών, η οποία στοχεύει στο να:

- Καταστήσει δυνατό ένα υψηλότερο επίπεδο ηλεκτρονικής επιχειρηματικότητας με την επιτάχυνση της μετάβασης σε ένα συνεπές σύνολο από πρότυπα διαχείρισης ταυτότητας
- Μειώσει την πολυπλοκότητα της ενοποίησης των επιχειρηματικών εφαρμογών
- Διαχειριστεί τη ροή χρηστών που εισέρχονται, χρησιμοποιούν και εγκαταλείπουν τον οργανισμό
- Υποστηρίξει σφαρικές προσεγγίσεις και σχήματα για συγκεκριμένες κατηγορίες λειτουργικών έργων
- Ανταποκρίνεται στην πίεση από τον διαρκώς αυξανόμενο αριθμό διαδικτυακών επιχειρηματικών εφαρμογών που χρειάζονται μεγαλύτερη ενοποίηση για δραστηριότητες όπως το single sign-on.

Η διαχείριση ταυτότητας εμπεριέχει δύο διαστάσεις: τη διάσταση της Ταυτότητας και τη διάσταση της Ταυτοποίησης [56]. Οι δύο αυτές διαστάσεις ορίζονται ως εξής:

- *Διάσταση Ταυτότητας*: το σύνολο των χαρακτηριστικών που αναπαριστούν ένα άτομο
- *Διάσταση Ταυτοποίησης*: το σύνολο των όρων, εννοιών και μηχανισμών που σχετίζονται με την αποκάλυψη αυτής της πληροφορίας και την χρήση αυτής.

Αυτή η διάκριση μεταξύ των δύο διαστάσεων έχει εισαχθεί προκειμένου να γίνει διάκριση σε δύο διαφορετικές (συμπληρωματικές) όψεις:

- Μια περιγραφική όψη η οποία αναφέρεται στην αναπαράσταση ενός ατόμου ή αντικειμένου αναφορικά με ένα σύνολο από σχετικά γνωρίσματα (όψη τρίτου προσώπου; αντικειμενοποίηση). Στην περίπτωση αυτή, η επίνοια της ταυτότητας λαμβάνει χώρα μέσω του προσδιορισμού ενός συνόλου γνωρισμάτων και σχετιζόμενων καταστάσεων που περιγράφουν τα χαρακτηριστικά ενός αντικειμένου (άτομα, ομάδες, οργανισμοί) το οποίο έχει μια ταυτότητα.
- Μια διαδικαστική όψη η οποία αναφέρεται στην ταυτοποίηση ενός ατόμου ή αντικειμένου μέσω της μοναδικής του διαφοροποίησης από όλα τα άλλα άτομα ή/και αντικείμενα (όψη τρίτου προσώπου; αντικειμενοποίηση), και μέσω της χρήσης αυτής της πληροφορίας. Στην περίπτωση αυτή, η επίνοια της ταυτότητας θεωρείται στα πλαίσια διαδικασιών σχετιζόμενων με την ταυτότητα με τις οποίες αντικείμενα (άτομα, ομάδες, οργανισμοί) που έχουν κάποια ταυτότητα σχετίζονται, όπως η αποκάλυψη της πληροφορίας της ταυτότητας (πιστοποίηση, σύνδεση με ένα προφίλ, κλπ) και ο τρόπος με τον οποίο η πληροφορία αυτή θα χρησιμοποιηθεί (για παροχή πρόσβασης σε πόρους, για παρακολούθηση και χαρακτηρισμό συμπεριφορών, κλπ).

Σε αυτό το σημείο είναι σημαντικό να τονιστεί ότι οι όροι Ταυτότητα (Identity) και Ταυτοποίηση (Identification) αναφέρονται σε δύο διαφορετικές έννοιες οι οποίες σχετίζονται αλλά δε θα πρέπει να συγχέονται. Από τη μια μεριά, η Ταυτότητα χρησιμοποιείται για να αναφερθεί σε ένα σύνολο από σαφή σχετικά γνωρίσματα (μόνιμα ή προσωρινά) ενός ατόμου στα πλαίσια πρακτικών δραστηριοτήτων. Για παράδειγμα, γνωρίσματα τα οποία σχετίζονται με την ικανότητα ενός ατόμου θα παρατίθενται στο εργασιακό πλαίσιο, σε ένα σενάριο στο οποίο η

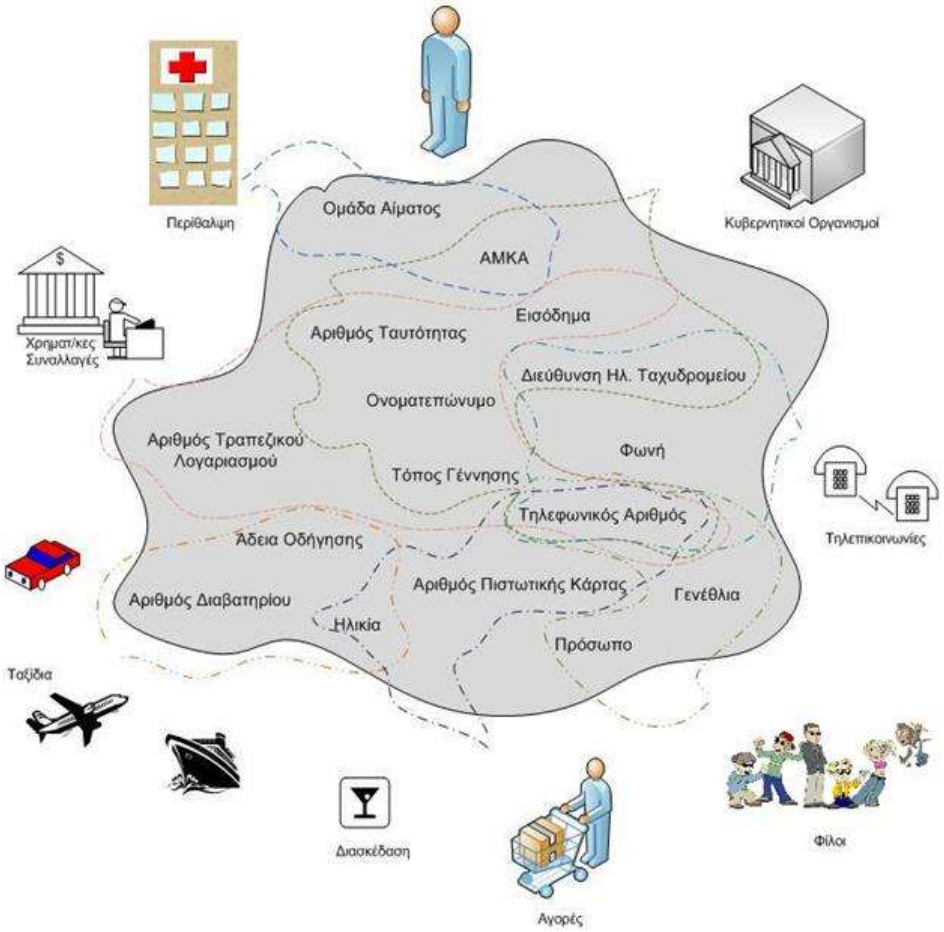
ικανότητα αποτελεί έναν σημαντικό παράγοντα επιτυχίας για την επιτυχή ολοκλήρωση ενός στόχου. Από την άλλη μεριά, η Ταυτοποίηση αναφέρεται στη διαδικασία που χρησιμοποείται για να συνδεθεί ένα άτομο με μια ταυτότητα. Κάποια κριτήρια μπορεί να χρησιμοποιηθούν για τον σκοπό αυτόν όπως το όνομα του ατόμου, τα δαχτυλικά του αποτυπώματα, τα γενετικά του χαρακτηριστικά και τα συμπεριφορικά του πρότυπα. Αυτή η μείωση των γνωρισμάτων είναι απαραίτητη για να παραχθούν εύκολα, αποδοτικά και αποτελεσματικά μέσα που επιτρέπουν στους ανθρώπους πρόσβαση σε περιορισμένη πληροφορία ή εύρεση ένος συγκεκριμένου ατόμου για λόγους ασφαλείας ή εμπορικούς λόγους, για παράδειγμα.

Μια ακόμη διάκριση που δε θα πρέπει να παραβλεφτεί αφορά στον αναγωγικό χαρακτηρισμό ενός ατόμου και στην ταυτότητα ενός ατόμου εν ζωή [5] [6]. Το πρώτο, γνωστό ως idem-identity, είναι στατικό ακόμη και αν τακτικά αναβαθμίζεται, και είναι το μόνο σαφώς επισημοποιημένο και χειριζόμενο από τεχνολογίες πληροφορίας και ταυτοποίησης. Το δεύτερο, γνωστό ως ipse-identity, αναπαριστά ποιο το άτομο πραγματικά είναι (από φιλοσοφική άποψη), είναι πρωταρχικώς ρευστό και απροσδιόριστο και βρίσκεται εκτός των ορίων των τεχνολογιών πληροφορίας και ταυτοποίησης.

2.2 *Τα ζητήματα Ταυτότητας και Ταυτοποίησης*

Η Ταυτότητα είναι ένα θέμα που δεν είναι περιορισμένο μόνο σε μια μικρή ομάδα ειδικών. Αντιθέτως, απαντάται σε πολλές όψεις της ζωής των ανθρώπων: στη γεωγραφική τους κινητικότητα (σε σχέση με τη διάσχιση περιοχών), στην ιδιωτική τους ζωή (σχετιζόμενη με τις δραστηριότητές τους, την ερωτική τους ζωή, κλπ), στην οικογενειακή τους ζωή (σε σχέση με την οικογενειακή τους κατάσταση και τη δομή της οικογένειάς τους), στην κοινωνική τους ζωή (αναφορικά με τους φίλους τους και τη σχέση τους με ομάδες), στην εργασιακή τους ζωή (σε

σχέση με τον ρόλο τους, τη θέση τους και τις ευθύνες τους) και στον τρόπο με τον οποίο πραγματοποιούν επιχειρηματικές δραστηριότητες (όπως η φήμη τους, κλπ), στη ζωή τους ως πολίτες (σε σχέση με την ψήφο τους και τη συμμετοχή τους στη ζωή της κοινότητας), στη βιολογική τους ζωή (όσον αφορά στην παροχή υπηρεσιών υγείας), στη ζωή τους ως πελάτες (σε σχέση με αγορές και συναλλαγές) μεταξύ των άλλων.



Εικόνα 1 Η ταυτότητα και διαφορετικές υπο-ταυτότητες ανά εφαρμογή

Στην πράξη, σε κάθε μία από αυτές τις διαφορετικές εκφάνσεις της ζωής, θέματα ταυτότητας και ταυτοποίησης μπορεί να προκύψουν και να λάβουν διαφορετικές μορφές. Έτσι, θέματα ταυτότητας και ταυτοποίησης μπορεί να σχετίζονται με τη νομιμότητα μιας πράξης δεδομένης της σχέσης του ατόμου με μια συγκεκριμένη ομάδα (χώρα, εταιρεία, κοινωνική ομάδα) ή λόγω

δεδομένων προνομίων. Για παράδειγμα, η ιθαγένεια μπορεί να προσφέρει πρόσβαση σε κάποια κοινωνικά οφέλη ή το δικαίωμα για ταξίδια και εργασία σε κάποια άλλη χώρα, ένα δίπλωμα ή κάποια άλλη απόδειξη δεξιοτήτων και ικανοτήτων μπορεί να επιτρέψει την αίτηση για μια θέση εργασίας και την κατοπινή εξάσκηση του επαγγέλματος, η φιλία μπορεί να προσφέρει τη δυνατότητα να ζητήσει και να λάβει δωρεάν υπηρεσίες από ένα άλλο άτομο (τον φίλο). Συνακολούθως, καθώς τα άτομα αναλαμβάνουν διαφορετικούς ρόλους κατά τη διάρκεια της ζωής τους, διαφορετικά σύνολα από χαρακτηριστικά που αφορούν σε αυτούς τους διαφορετικούς ρόλους χρησιμοποιούνται για την αναπαράσταση της ταυτότητας τους. Κάθε μία από αυτές τις «μερικές ταυτότητες» περιλαμβάνει τόσο κληρονομήσιμα «μόνιμα» χαρακτηριστικά (όπως η εθνικότητα, το φύλο, κλπ) και χαρακτηριστικά τα οποία απέκτησαν κατά τη διάρκεια της ζωής τους (όπως δίπλωμα, ικανότητες, κλπ) ή τα οποία τους έχουν ανατεθεί ή εκδοθεί προκειμένου να προχωρήσουν με τον ρόλο τους (όπως μια θέση, κάποιου είδους αρμοδιότητα, κλπ).

Μια άλλη διάσταση σχετίζεται με την αποτελεσματική απόδειξη (με διαφορετικά επίπεδα αξιοπιστίας) ότι ένα άτομο έχει πραγματικά την άδεια ή τα διαπιστευτήρια που ισχυρίζται και τα οποία απαιτούνται για μια πράξη. Παραδείγματα τέτοιων στοιχείων είναι μια ταυτότητα (διαβατήριο ή επαγγελματική κάρτα), ένα κλειδί (που αποδεικνύει σε μια τεχνική υποδομή το δικαίωμα πρόσβασης), ένα δίπλωμα, ένα κοινωνικό αποδεικτικό ή μια σύσταση (για παράδειγμα από έναν γνώριμο).

Άλλες όψεις σχετίζονται με τη (μερική) πρόσβαση σε αυτήν την πληροφορία ταυτότητας από άλλους, τη χρήση αυτής της πληροφορίας και το ζήτημα του ελέγχου (για παράδειγμα στο [7] παρατίθενται συζητήσεις για τον έλεγχο ανωνυμίας). Η διαχείριση της πρόσβασης στην πληροφορία και του ελέγχου (από το άτομο, θεσμικά όργανα, οργανισμούς, εμπορικές

οντότητες) είναι κρίσιμη καθώς σχετίζεται με την ελευθερία δράσης ενός ατόμου. Για παράδειγμα, η αποκάλυψη πληροφορίας σχετικής με τις πολιτικές πεποιθήσεις ενός ατόμου μπορεί να έχει σημαντικές επιπτώσεις στους βαθμούς ελευθερίας δράσης του ατόμου αυτού (στη «χειρότερη περίπτωση» το άτομο αυτό μπορεί να βρεθεί στη φυλακή, σε άλλες περιπτώσεις μπορεί να διακινδυνεύσει την εργασιακή του θέση). Συγκεκριμένα, η μεγάλη διαφάνεια της πληροφορίας μπορεί να οδηγήσει τους ανθρώπους σε πλήρη απραξία υπό τον φόβο των αντίποινων από άλλα άτομα, ομάδες ή την κοινωνία, οδηγώντας σε σημαντικές αρνητικές επιπτώσεις (άτομα μπορεί να φοβούνται να καταγγέλουν μη αποδεκτές καταστάσεις) ή θετικές (αποτροπή των ατόμων από την απόκρυψη κερδών και πληρωμή λιγότερων φόρων ή απόδοση ευθυνών στα σωστά άτομα για την πρόκληση μιας ζημιάς). Μια πιο κοινότυπη όψη σχετίζεται με την ανερυθρίαστη εκμετάλλευση αυτής της πληροφορίας από τρίτους οι οποίοι τη θεωρούν κοινό αγαθό, με ένα από τα πιο γνωστά παραδείγματα αυτής να είναι η άμεση προώθηση μαζικής ηλεκτρονικής αλληλογραφίας (spamming).

2.3 *Η έννοια της Ταυτότητας*

Η έννοια της Ταυτότητας σχετίζεται με τον χαρακτηρισμό και την αναπαράσταση ενός ατόμου (φυσικού ή ηθικού) ή μιας ομάδας, και αφορά στη δομή αυτού του χαρακτηρισμού. Για παράδειγμα, η Ταυτότητα μπορεί να κατηγοριοποιηθεί σύμφωνα με τις διαφορετικές εκφάνσεις, όπως η προσωπική Ταυτότητα (προσωπική), η βιολογική Ταυτότητα (DNA - Deoxyribonucleic acid), η κοινωνική Ταυτότητα (συμμετοχή) ή η νομική Ταυτότητα και τις συναθροίζει με βάση τη χρήση σε διαφορετικές καταστάσεις (όπως δραστηριότητες ελεύθερου χρόνου, συνναλαγές, εργασιακές ή κοινωνικές αλληλεπιδράσεις). Η έννοια της Ταυτότητας μπορεί να εφαρμοσθεί σε ένα φυσικό, ηθικό ή αφηρημένο άτομο (όπως ένας οργανισμός ή ομάδα).

Μια δεύτερη κατηγοριοποίηση της ταυτότητας παρουσιάζεται στα Τρία Στρώματα Ταυτότητας [8]. Στο μοντέλο αυτό, ο Durand διακρίνει τρεις κατηγορίες (ή στρώματα) ταυτοτήτων:

- Η *προσωπική ταυτότητα* (η εσωτερική και παντοτινή ταυτότητα). Αυτή είναι η αληθινή προσωπική ταυτότητα που ανήκει και ελέγχεται αποκλειστικά από το άτομο.
- Η *συλλογική ταυτότητα* (η ανατεθημένη ταυτότητα). Η ταυτότητα αυτή σχετίζεται με ένα συγκεκριμένο πλαίσιο (για παράδειγμα μια επιχειρηματική σχέση) και αναπαριστά ένα προσωρινά ανατεθιμένο και εκδωσμένο χαρακτηριστικό για το άτομο όπως ο εργασιακός του τίτλος, το τηλεφωνικό του νούμερο, κλπ.
- Η *προωθητική ταυτότητα* (η αφηρημένη ή ενοποιημένη ταυτότητα). Αυτή η ταυτότητα αφορά στο αποτέλεσμα ανάλυσης του προφίλ του ατόμου. Το άτομο δε θεωρείται πραγματικά ως ένα πρόσωπο (το άτομο δεν έχει όνομα), αλλά ως το αποτέλεσμα φιλτραρίσματος που λαμβάνει χώρα επί ενός συνόλου χαρακτηριστικών. Ένα παράδειγμα θα μπορούσε να είναι ο πελάτης ο οποίος είναι νεαρής ηλικίας, έχει πανεπιστημιακή εκπαίδευση, είναι κάτοχος μοτοσυκλέτας παλαιότητας μεγαλύτερης των τριών ετών και παίζει τένις, με τον οποίο επικοινωνεί ο πωλητής.

Ενώ το μοντέλο αυτό μπορεί να φαίνεται πολύ απλουστευμένο για να ενσωματώνει όλη την πολυπλοκότητα της έννοιας της Ταυτότητας, εισάγει βασικές ιδιότητες στην Ταυτότητα και πιο συγκεκριμένα την ιδιότητα του *προσωρινού*, την ιδιότητα της *εξάρτησης* και την ιδιότητα του *συγκεκριμένου*.

2.4 *Η έννοια της Ταυτοποίησης*

Στην προηγούμενη παράγραφο παρουσιάστηκε η περιγραφική όψη της Ταυτότητας, δηλαδή μια εννοιολογικοποίηση που βασίζεται σε ένα σύνολο από ξεχωριστά χαρακτηριστικά τα οποία

ανήκουν σε ένα αντικείμενο και τα οποία αποτελούν την ταυτότητά του. Στην τρέχουσα παράγραφο θα παρουσιαστεί η διαδικαστική όψη της Ταυτότητας, με άλλα λόγια διαδικασίες με τις οποίες η ταυτότητα ενός αντικειμένου συνδέεται και συγκεκριμένα πώς η πληροφορία της ταυτότητας αποκαλύπτεται και χρησιμοποιείται. Έτσι, καλύπτει πλήθος εννοιών όπως η ανωνυμία, η μη συνδεσιμότητα και οι ταυτοποιητές, τις οποίες θα δούμε σε επόμενες παραγράφους.

Η ταυτοποίηση είναι ως επί τω πλείστον μερική (αποκάλυψη μόνο μιας μερικής ταυτότητας) και χρησιμοποιείται σε ένα συγκεκριμένο πλαίσιο και για έναν συγκεκριμένο σκοπό, όπως είναι η εκχώρηση πρόσβασης σε έναν πόρο ή μια λειτουργία. Θα πρέπει να σημειωθεί, όπως θα δούμε και παρακάτω, ότι και η σύνδεση με προφίλ (profiling) αντιπροσωπεύει μια έννοια η οποία μπορεί να σχετίζεται με την ταυτοποίηση. Ο όρος αυτός αφορά στη διαδικασία κατασκευής ή εφαρμογής ενός προφίλ ενός ατόμου ή μιας ομάδας. Ένα προφίλ αποτελείται από πρότυπα συσχετισμένων δεδομένων [9].

2.4.1 Η χρησιμότητα της ταυτοποίησης

Τα κύρια πλαίσια εντός των οποίων η ταυτοποίηση βρίσκει χρησιμότητα είναι τα ακόλουθα:

- Έλεγχος πρόσβασης σε περιορισμένους πόρους ή περιοχές
- Εκμετάλλευση της πληροφορίας ταυτότητας
- Παρακολούθηση για απόδοση ευθυνών

2.4.1.1 Έλεγχος πρόσβασης σε περιορισμένους πόρους ή περιοχές (πιστοποίηση)

Ο έλεγχος αυτός αφορά τόσο στην πιστοποίηση όσο και στη διαχείριση πρόσβασης. Η πιστοποίηση σχετίζεται με την επιβεβαίωση της ταυτότητας του ατόμου και συγκεκριμένα με τη διασφάλιση ότι το άτομο είναι όντως αυτό που ισχυρίζεται ότι είναι. Η διαχείριση του

δικαιώματος πρόσβασης σχετίζεται με τον ρόλο του ατόμου εντός ενός συγκεκριμένου πλαισίου και τις κατηγορίες των λειτουργιών που έχουν ανατεθεί και οι οποίες επιτρέπονται στο άτομο (για παράδειγμα μπορεί να υφίσταται κάποιος περιορισμός σχετικά με τη χρήση ενός πόρου ή τις επιτρεπόμενες λειτουργίες σε μια συγκεκριμένη περιοχή).

Από την οπτική γωνία του χρήστη, η ταυτοποίηση μπορεί να εφαρμοστεί σε ένα άλλο άτομο ή έναν οργανισμό. Στοιχεία ταυτότητας μπορούν επίσης να χρησιμοποιηθούν από ένα άτομο προκειμένου να διασφαλίσει ότι ένα άλλο άτομο ή οργανισμός είναι όντως αυτό το οποίο ισχυρίζεται ότι είναι. Έτσι, για παράδειγμα, μια διεύθυνση ηλεκτρονικού ταχυδρομείου παρέχει κάποιου βαθμού ταυτοποίηση σχετιζόμενη με τον αποστολέα, ενώ ο Ενιαίος Εντοπιστής Πόρων (Unified Resource Locator - URL) μπορεί να παρέχει κάποια μέσα για την πιστοποίηση ενός οργανισμού.

2.4.1.2 Αναγνώριση με σκοπό την εκμετάλλευση της πληροφορίας ταυτότητας (γνώση)

Ένας επίσης σημαντικός λόγος για την ταυτοποίηση ενός ατόμου είναι να επιτρέψει την πρόσβαση σε σχετική πληροφορία η οποία μπορεί να είναι απαραίτητη για το σύστημα, αλλά και να προσδώσει μια σειρά από πλεονεκτήματα στο άτομο, όπως για παράδειγμα επιτρέποντας πιο προσαρμοσμένη και αποτελεσματική αλληλεπίδραση μεταξύ του συστήματος και του ατόμου. Σε άλλες περιπτώσεις, η πρόσβαση σε αυτήν την πληροφορία αποτελεί υποχρεωτική συνθήκη για την παράδοση μιας συγκεκριμένης υπηρεσίας.

Αντιστρόφως, η πρόσβαση στην πληροφορία αυτή μπορεί να αποδειχθεί επιζήμια για το άτομο. Για παράδειγμα, εμπορικές εταιρείες άμεσης προώθησης μπορεί να εκμεταλευτούν την πληροφορία αυτήν προκειμένου να χειριστούν πιο αποτελεσματικά προς όφελός τους τον τελικό καταναλωτή στον οποίο στοχεύουν, με την πληροφορία αυτή να προσφέρει ένα άδικο πλεονέκτημα έναντι του ατόμου κατά τη διάρκεια μιας διαπραγμάτευσης.

2.4.1.3 Παρακολούθηση και Απόδοση Ευθυνών

Η παρακολούθηση και η απόδοση ευθυνών σχετίζονται με τη δυνατότητα καταγραφής και ελέγχου των πράξεων ενός ατόμου (και τη σύνδεσή του με μια μερική ταυτότητα) και μπορούν να χρησιμοποιηθούν εντός πλήθους πλαισίων. Στον εμπορικό κόσμο, είναι δυνατό να βοηθήσουν υποστηρικτικά σε διάφορες φάσεις μιας συναλλαγής, όπως είναι οι πληρωμές. Σημαντική είναι η υποστήριξη που μπορούν να προσφέρουν στις κοινωνικές όψεις των ηλεκτρονικών εργαλείων αγορών, συνεισφέροντας για παράδειγμα, στη διαμόρφωση της φήμης των εμπόρων ή των πελατών, με χαρακτηριστικό παράδειγμα το eBay.

Σε μια επικοινωνιακή δραστηριότητα, η παρακολούθηση αυτή μπορεί να περιλαμβάνει την παροχή πληροφορίας (όπως είναι το όνομα πρόσβασης, η IP διεύθυνση, κλπ) η οποία μπορεί να χρησιμοποιηθεί από τον συγγραφέα (ή τη νοητή ταυτότητά του), ενώ στον χώρο της ασφάλειας, πληροφορία μπορεί να καταγράφεται προκειμένου να αναγνωριστούν ύποπτες δραστηριότητες, όπως για παράδειγμα δραστηριότητες ξεπλύματος χρήματος.

2.4.2 Οι κίνδυνοι της Ταυτοποίησης

Είναι σημαντικό να αναφερθεί ότι η Ταυτοποίηση, ανεξαρτήτως της χρήσης της, μπορεί να συνοδεύεται από σημαντικές αρνητικές επιπτώσεις για το άτομο είτε λόγω λανθασμένης ταυτοποίησης είτε λόγω μη επιθυμητής ταυτοποίησης.

2.4.2.1 Λανθασμένη Ταυτοποίηση

Οι περισσότερες περιπτώσεις λανθασμένης ταυτοποίησης αφορούν σε σφάλματα των συστημάτων αναγνώρισης ή σε επιτυχημένες προσπάθειες *κλοπής ή πλαστογράφησης ταυτότητας* (*identity theft – identity fraud*). Η πρώτη αιτία των σφαλμάτων οφείλεται στις εγγενείς αδυναμίες

των συστημάτων αναγνώρισης σε τεχνικό επίπεδο, σε σφάλματα κατά την καταχώρηση των δεδομένων ή σε απροσεξία σε επίπεδο ανθρώπου. Εστιάζοντας στις περιπτώσεις κλεμμένης ή πλαστής ταυτότητας, οι επιπτώσεις μιας τέτοιας ενέργειας μπορεί να έχει κυμαινόμενες επιπτώσεις αναλόγως με την εφαρμογή και το είδος της πληροφορίας η οποία χρησιμοποιείται για την ταυτοποίηση αλλά και την πληροφορία η οποία συνδέεται με την ταυτοποίηση αυτή. Από τις πιο σοβαρές επιπτώσεις αυτών είναι η αποκάλυψη εμπιστευτικής πληροφορίας σε τρίτους ή η απώλεια μεγάλων χρηματικών ποσών (μέσω για παράδειγμα αν μια επιχείρηση εξαπάτησης πείσει ένα άτομο να αποκαλύψει τις πληροφορίες της πιστωτικής του κάρτας). Παραδείγματα κλοπής ταυτότητας αποτελούν η κλοπή των κωδικών πρόσβασης και η κλοπή εγγράφων ταυτότητας μεταξύ των άλλων. Περιπτώσεις απάτης ταυτότητας αποτελούν η πανομοιότυπη αναπαραγωγή βιομετρικών δεδομένων και το μάντεμα των κωδικών πρόσβασης.

2.4.2.2 Μη επιθυμητή Ταυτοποίηση

Η μη επιθημητή ταυτοποίηση αφορά στις περιπτώσεις κατά τις οποίες πληροφορία σχετική με το άτομο αποκαλύπτεται σε τρίτους χωρίς τη συναίνεσή του ή/και τη γνώση του. Για τον λόγο αυτόν, αυτός ο κίνδυνος είναι άμεσα συνδεδεμένος με την *ιδιωτικότητα* του ατόμου. Παραδείγματα απαντώνται σε διάφορους τομείς της ανθρώπινης δραστηριότητας. Έτσι στις εμπορικές συναλλαγές, και κυρίως στις ηλεκτρονικές, η αποκάλυψη πληροφορίας σε τρίτους για το άτομο χωρίς τη συναίνεσή του μπορεί να χρησιμοποιηθεί για την ενίσχυση της στοχευμένης προώθησης ενός προϊόντος μέσω της χειραγώγησής του. Στον εργασιακό χώρο, πληροφορία για το άτομο σχετικά με τις πολιτικές του πεποιθήσεις, τη συμμετοχή του σε μια οργάνωση, το ιατρικό του ιστορικό μεταξύ των άλλων μπορεί να επιφέρει σημαντικές επιπτώσεις στο άτομο, συμπεριλαμβανομένων της εργασιακής πίεσης, της διάκρισης, ακόμη και της απώλειας της εργασίας. Όσον αφορά στον κοινωνικό χώρο, τόσο η διαρροή των πολιτικών πεποιθήσεων του

ατόμου όσο και των σεξουαλικών του προτιμήσεων μπορεί να έχει επίσης σημαντικό αρνητικό αντίκτυπο.

Είναι σημαντικό σε αυτό το σημείο να τονίσουμε ότι αυτού του είδους ο κίνδυνος ταυτοποίησης προκύπτει από την αποκάλυψη τόσο των στοιχείων τα οποία είναι άμεσα συνδεδεμένα με τις λειτουργίες του συστήματος το οποίο και την πραγματοποιεί όσο και στοιχείων από άλλα συστήματα και βάσεις δεδομένων με τα οποία επιτρέπει σύνδεση. Η δεύτερη περίπτωση είναι εφικτή όταν το αναγνωριστικό για το άτομο είναι κοινό στα συστήματα τα οποία μοιράζονται πληροφορίες, όπως Αριθμός Φορολογικού Μητρώου (ΑΦΜ), Αριθμός Δελτίου Ταυτότητας (ΑΔΤ), Αριθμός Μητρώου Κοινωνικής Ασφάλισης (ΑΜΚΑ). Η νομιμότητα και αποδοχή τέτοιων πράξεων εξαρτάται από την αναγκαιότητα αυτών καθώς και από το νομικό υπόβαθρο το οποίο μπορεί να τις στηρίζει ή να τις καταδικάζει. Έτσι, για παράδειγμα η Τράπεζα της Ελλάδος στην προσπάθειά της για πρόληψη και καταστολή της νομιμοποίησης εσόδων από εγκληματικές δραστηριότητες και της χρηματοδότησης της τρομοκρατίας στα πλαίσια της 3^{ης} Κοινοτικής Οδηγίας (3rd Anti-Money Laundering Directive) καθορίζει μια σειρά ενεργειών για τους Χρηματοπιστωτικούς Οργανισμούς και τα Πιστωτικά Ιδρύματα η οποία προϋποθέτει τη συνεργασία των τελευταίων προκειμένου να εντοπίζουν, προλαμβάνουν, αποτρέπουν και αναφέρουν συναλλαγές τέτοιου είδους [53].

2.4.3 Διαδικασίες Ταυτοποίησης

Με βάση τα παραπάνω διακρίνουμε δύο προσεγγίσεις για την ταυτοποίηση ενός ατόμου, τη *φανερή* (explicit) και την *υποκρυπτόμενη* (implicit) ταυτοποίηση [56]. Όπως φανερώνουν και οι όροι, η πρώτη περίπτωση αφορά στην ταυτοποίηση του ατόμου για την οποία το άτομο έχει γνώση και ίσως ακόμη και να συμμετέχει σε αυτήν, ενώ η δεύτερη αφορά στην ταυτοποίηση του ατόμου δίχως το τελευταίο να έχει γνώση και να έχει δώσει τη συναίνεσή του. Η εν γνώση

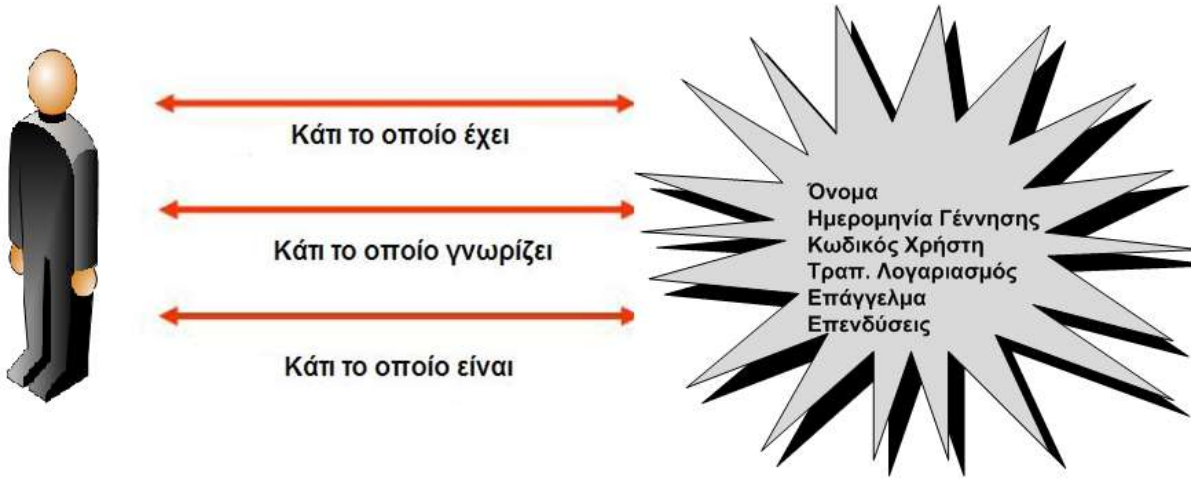
ταυτοποίηση του ατόμου μπορεί να λάβει χώρα μέσω της παρουσίασης εκ μέρους του ατόμου ή ενός εξουσιοδοτημένου ατόμου του δελτίου ταυτότητας, της επαγγελματικής του κάρτας, της πιστωτικής του κάρτας ή των βιομετρικών του χαρακτηριστικών μεταξύ των άλλων. Επίσης, στην κατηγορία αυτή εμπίπτουν οι μηχανισμοί συλλογής των δεδομένων ταυτότητας του ατόμου κατά τη φάση αρχικής καταχώρησης η οποία λαμβάνει χώρα με τη συναίνεση του ατόμου.

Στην περίπτωση της υποκρυπτόμενης ταυτοποίησης, η οποία και είναι άμεσα συνδεδεμένη με τους κινδύνους μη επιθυμητής ταυτοποίησης οι οποίοι παρουσιάστηκαν στην προηγούμενη παράγραφο, τόσο η διαδικασία ταυτοποίησης όσο και η λήψη της πληροφορίας ταυτότητας λαμβάνουν χώρα χωρίς τη συγκατάθεση ή/και τη γνώση του ατόμου. Έτσι, για παράδειγμα, με βάση αναγνωριστικά του ατόμου, όπως είναι οι κωδικοί του, η φυσική του παρουσία σε έναν ή περισσότερους χώρους, είναι δυνατή η συλλογή των ιχνών της δράσης του ατόμου αυτού και η εξαγωγή συμπερασμάτων και πληροφοριών για την ταυτότητα και τη δραστηριότητα αυτού εν αγνοία του.

2.4.4 Χαρακτηριστικά προς ταυτοποίηση

Η διαδικασία ταυτοποίησης ενός ατόμου μπορεί να βασίζεται σε πληθώρα χαρακτηριστικών του τα οποία γενικώς διακρίνονται σε :

- Κάτι που το άτομο γνωρίζει: αφορά σε πληροφορία την οποία το άτομο γνωρίζει, όπως είναι ο προσωπικός του κωδικός για την πρόσβαση σε ένα σύστημα,
- Κατι που το άτομο έχει: περιλαμβάνει περιπτώσεις κατά τις οποίες το άτομο είναι κάτοχος ενός αντικειμένου αναγνωριστικού, όπως είναι μια κάρτα, ψηφιακά πιστοποιητικά, κλπ,
- Κάτι που το άτομο είναι ή κάνει: σχετίζεται με γνωρίσματα με τα οποία το άτομο είναι άμεσα συνδεδεμένα, με χαρακτηριστικό παράδειγμα τα βιομετρικά του χαρακτηριστικά.



Εικόνα 2 Χαρακτηριστικά προς ταυτοποίηση

Όλα αυτά τα χαρακτηριστικά διαφέρουν ως προς την πρακτικότητα τους, το κόστος εφαρμογής τους, το πεδίο εφαρμογής τους, την αξιοπιστία τους, την αποτελεσματικότητα τους, την ασφάλειά τους αλλά και τον σεβασμό τους προς την ιδιωτικότητα του ατόμου.

2.4.5 Το ζήτημα της ιδιωτικότητας

Όπως ήδη αναφέρθηκε, η ταυτοποίηση δεν είναι πάντα επιθυμητή, με αποτέλεσμα να τίθενται σοβαρά θέματα σχετικά με τη διασφάλιση της ιδιωτικότητας του ατόμου και συνακολούθως να καθίσταται απαραίτητη η μελέτη και η ανάπτυξη μηχανισμών οι οποίοι θα στοχεύουν στη διασφάλισή της. Η ανάγκη της προστασίας της ιδιωτικότητας γίνεται ακόμη πιο επιτακτική με τις σύγχρονες εξελίξεις στην τεχνολογία (όπως είναι η σύνθεση του προφίλ) οι οποίες μπορεί να κατθιστούν δυνατή τη χειραγώγηση των ατόμων σε ευρεία κλίμακα, να μειώνουν την ελευθερία αυτοπροσδιορισμού και προσωπικής αυτονομίας και να καταστρατηγούν την κοινωνική ελευθερία, καταλήγωντας να ασκούν τεράστια επίδραση ακόμη και στη δημοκρατία [9].

Ποικίλες έννοιες, μέσα και μηχανισμοί μπορούν να χρησιμοποιηθούν για την προστασία της ιδιωτικότητας του ατόμου [54], οι οποίες κυρίως «θολώνουν» τη διαδικασία της ταυτοποίησης, κρύβοντας χαρακτηριστικά ή ίχνη του ατόμου και παρουσιάζονται στις ακόλουθες παραγράφους.

2.4.5.1 Η έννοια της Μη Συνδεσιμότητας

Η έννοια της Μη Συνδεσιμότητας (Unlinkability) δύο ή περισσότερων πραγμάτων (όπως υποκείμενα, πράξεις) αφορά στο ότι «εντός του συστήματος τα πράγματα αυτά δεν είναι ούτε περισσότερο ούτε λιγότερο συσχετισμένα απ’ότι είναι συσχετισμένα με βάση την προγενέστερη γνώση (a-priori knowledge)» [54]. Ο όρος αυτός είναι γενικός και αφορά στη μη συνδεσιμότητα κάθε είδους πραγμάτων. Ένας ορισμός πιο εστιασμένος στον χρήστη παρέχεται [55] ως εξής: διασφαλίζει ότι ένας χρήστης μπορεί να κάνει πολλαπλή χρήση πόρων ή υπηρεσιών χωρίς τρίτοι να μπορούν να συνδέσουν αυτές τις χρήσεις μεταξύ τους. Η Μη Συνδεσιμότητα απαιτεί οι χρήστες και/ή τα υποκείμενα να μην είναι σε θέση να καθορίσουν αν ο ίδιος χρήστης προκάλεσε

συγκεκριμένες ενεέργειες στο σύστημα. Με άλλα λόγια, καμία μεταβολή της γνώσης σχετικά με τη σύνδεση μεταξύ χρήσεων δεν είναι εφικτή.

Η Μη Συνδεσιμότητα ενός πράγματος είναι δυνατό να είναι μερική και να «προστατεύει» μόνο κάποιες από τις δράσεις που συνδέονται με το άτομο αυτό. Έτσι, για παράδειγμα, η μη συνδεσιμότητα ενός πράγματος μπορεί να αφορά μόνο στη σύνδεσή του με την πηγή ενός μηνύματος (όπως είναι ο συγγραφέας του) ή τον παραλήπτη του.

2.4.5.2 Η έννοια της Μη Παρατηρησιμότητας

Ένας γενικός ορισμός της έννοιας της Μη Παρατηρησιμότητας (Unobservability) αφορά στην κατάσταση των αντικειμένων ενδιαφέροντος κατά την οποία αυτά δε μπορούν να διακριθούν από κάθε άλλο αντικείμενο ενδιαφέροντος [54]. Στο [55] απαντάται ένας λιγότερο γενικός ορισμός. Σύμφωνα με τον τελευταίο η Μη Παρατηρησιμότητα «διασφαλίζει ότι ένας χρήστης μπορεί να κάνει χρήση ενός πόρου ή μιας υπηρεσίας χωρίς άλλοι, και κυρίως τρίτοι, να είναι σε θέση να παρατηρήσουν ότι ο πόρος ή η υπηρεσία χρησιμοποιείται. [...] Η Μη Παρατηρησιμότητα απαιτεί χρήστες και/ή υποκείμενα να μη μπορούν να καθορίσουν αν μια λειτουργία λαμβάνει χώρα».

Με βάση τον πρώτο – πιο γενικό – ορισμό, η μέθοδος με την οποία μπορεί να επιτευχθεί η Μη Παρατηρησιμότητα βασίζεται στην εξασφάλιση της μη διακρισιμότητας των αντικειμένων ενδιαφέροντος. Η Μη Παρατηρησιμότητα είναι πιο δυνατή από τη Μη Συνδεσιμότητα καθώς προστατεύει το περιεχόμενο μιας λειτουργίας, ακόμη και την ίδια την ύπαρξη της τελευταίας. Σαφώς ένα μη παρατηρήσιμο αντικείμενο είναι και μη συνδέσιμο, αφού μια βασική προϋπόθεση της συνδεσιμότητας είναι η γνώση της ύπαρξης του αντικειμένου [56]. Παρόμοια έννοια αποτελεί η *μη ιχνηλασιμότητα* (untraceability), με τον ορισμό της να είναι η δυνατότητα να μην

εντοπιστεί επικοινωνία μεταξύ διαφορετικών στοιχείων εφαρμογής και συνακολούθως να συλλεγεί ιδιωτική πληροφορία.

2.4.5.3 Η έννοια της Ανωνυμίας

Με βάση την ίδια μελέτη [54] «η Ανωνυμία είναι η κατάσταση του να μην είναι ένα υποκείμενο αναγνωρίσιμο εντός ενός συνόλου υποκειμένων, το σύνολο ανωνυμίας (anonymity set). Το σύνολο ανωνυμίας είναι το σύνολο όλων των πιθανών υποκειμένων. Όσον αφορά στους διάφορους χρήστες, το σύνολο ανωνυμίας αποτελείται από τα υποκείμενα τα οποία μπορεί να προκαλέσουν μια πράξη».

Θα πρέπει να τονιστεί ότι ο ορισμός αυτός εφαρμόζεται σε κάθε είδους υποκείμενα και όχι μόνο σε χρήστες. Ένας ορισμός πιο εστιασμένος στους χρήστες [55] τονίζει ότι η Ανωνυμία «διασφαλίζει ότι ένας χρήστης μπορεί να χρησιμοποιεί έναν πόρο ή μια υπηρεσία χωρίς να αποκαλύπτει την ταυτότητά του. Οι απαιτήσεις για Ανωνυμία παρέχουν προστασία για την ταυτότητα του χρήστη. Η Ανωνυμία δεν έχει στόχο να προστατεύει την ταυτότητα του υποκειμένου. [...] Η Ανωνυμία απαιτεί οι άλλοι χρήστες να μην είναι σε θέση να καθορίσουν την ταυτότητα ενός χρήστη ο οποίος συνδέεται με ένα υποκείμενο ή με μια λειτουργία»

Μπορούν να διακριθούν διαφορετικά επίπεδα ελέγχου της ανωνυμίας [57] : άνευ όρων ανωνυμία, ελεγχόμενη από τον χρήστη υπό όρους ανωνυμία και ελεγχόμενη από έμπιστο τρίτο υπό όρους ανωνυμία. Η άνευ όρων ανωνυμία αφορά σε μη εφικτή ανάκληση της ανωνυμίας, δηλαδή αφορά σε μόνιμη, μη αναστρέψιμη, κατάσταση ανωνυμίας. Σε μερικές εφαρμογές, όμως, είναι πιθανό ο χρήστης να επιθυμεί να ανακαλέσει την ανωνυμία του, όπως για παράδειγμα, στην περίπτωση μιας ιατρικής βάσης δεδομένων, από την οποία ο ασθενής επιθυμεί να ζητήσει το ιατρικό του ιστορικό οπότε και θα πρέπει να αποδείξει την ταυτότητά του. Σε άλλες περιπτώσεις η ανάκληση της ανωνυμίας μπορεί να είναι εφικτή από τρίτους υπό

συγκεκριμένες συνθήκες, όπως για παράδειγμα για την καταπολέμηση εγκληματικών δραστηριοτήτων.

2.4.5.4 Η έννοια της Ψευδωνυμίας

Τα ψευδώνυμα (pseudonyms) στην περίπτωσή μας αποτελούν αναγνωριστικά υποκειμένων (αποστολέων και παραληπτών) [54]. Το υποκείμενο στο οποίο αναφέρεται το ψευδώνυμο είναι ο κάτοχος αυτού. Ο όρος αυτός, προερχόμενος από την αντίστοιχη ελληνική λέξη, σημαίνει «λανθασμένως ονομασμένος». Έτσι, ουσιαστικά σημαίνει ένα όνομα διαφορετικό από το πραγματικό. Καθώς το πραγματικό όνομα (όπως αυτό δηλώνεται στα επίσημα έγγραφα ταυτότητας τα οποία έχουν εκδοθεί από το κράτος) είναι αυθαίρετο (μπορεί ακόμη και να αλλάξει κατά τη διάρκεια της ζωής του ατόμου, έγινε μια προέκταση του όρου «ψευδώνυμο» [56] ώστε να συμπεριλάβει όλα τα αναγνωριστικά, ακόμη και όλα τα ονόματα.

Τα ψευδώνυμα αποτελούν έναν ιδιαίτερο έμμεσο μηχανισμό ο οποίος βοηθά στην απομόνωση και στην προστασία της ταυτότητας του ατόμου όταν το τελευταίο πραγματοποιεί κάποια δραστηριότητα. Η διάκρισή τους από την ανωνυμία είναι σαφής, καθώς τα ψευδώνυμα αποτελούν χαρακτηριστικά, μπορεί να έχουν μονιμότητα και επιτρέπουν συνδεσιμότητα. Η διαδικασία της ψευδωνυμίας είναι γνωστή και ως απο-ταυτοποίηση (de-identification). Χαρακτηριστικό παράδειγμα της αναγκαιότητας ψευδωνυμίας απαντάται στην ιατρική έρευνα. Στον χώρο αυτόν, ένα μέρος των δεδομένων είναι ιδιαιτέρως ευαίσθητα και απαιτούν υψηλό βαθμό προστασίας. Παρ' όλα αυτά η ανωνυμία δεν αποτελεί μια καλή λύση σε πολλές περιπτώσεις, καθώς απαγορεύει τη συνδεσιμότητα, η οποία κρίνεται απαραίτητη σε πληθώρα καταστάσεων. Έτσι, σε πολλές περιπτώσεις, ένας ασθενής απαιτείται να μπορεί να ταυτοποιηθεί προκειμένου να είναι εφικτή η παρακολούθηση της πορεία της ασθένειας από την οποία υποφέρει και του τρόπου με τον οποίο αντιδρά στην παρεχόμενη σε αυτόν θεραπεία. Δεδομένης

της διάσπαρτης πληροφορίας για έναν ασθενή σε πλήθος συστημάτων και πόρων αλλά και της αναγκαιότητας για προστασία της ιδιωτικότητάς του, η ψευδωνυμία αποτελεί μια αποτελεσματική μέθοδο.

2.4.5.5 Η έννοια της Κωδικοποίησης

Ο όρος αυτός είναι συσχετισμένος με την προστασία του αντικειμένου ενδιαφέροντος, σε αντίθεση με τη μη παρατηρησιμότητα η οποία σχετίζεται με την προστασία των διαδικασιών στις οποίες συμμετέχει ή τις οποίες πραγματοποιεί το αντικείμενο αυτό [56]. Με άλλα λόγια, η κυκλοφορία ενός «κωδικοποιημένου» αντικειμένου μπορεί να είναι παρατηρήσιμη, και ο παραλήπτης και ο αποστολέας αυτού του αντικειμένου να είναι ορατώς συνδεδεμένοι μεταξύ τους, αλλά η φύση αυτού του συνδέσμου και το περιεχόμενο του αντικειμένου θα παραμείνουν άγνωστα και ιδιωτικά. Χαρακτηριστικό παράδειγμα αποτελεί η κωδικοποίηση του περιεχομένου ενός ηλεκτρονικού γράμματος η οποία στοχεύει στην προστασία της ταυτότητας του ατόμου μέσω της απόκρυψης της πληροφορίας του περιεχομένου.

3 *Βιομετρικά Συστήματα*

Στο κεφάλαιο αυτό αναλύονται τα βιομετρικά συστήματα ως συστήματα ταυτοποίησης, επιβεβαίωσης ταυτότητας αλλά και σύνθεσης του προφίλ ατόμων. Συγκεκριμένα, μετά από μια σύντομη ιστορική αναδρομή η οποία στοχεύει στην ανάδειξη της διαχρονικότητας της κύριας ιδέας η οποία κρύβεται πίσω από τα βιομετρικά συστήματα σχετικά με τη χρήση βιομετρικών χαρακτηριστικών ως ανθρώπινων ταυτοποιητών, παρουσιάζονται οι διάφορες κατηγοριοποιήσεις των βιομετρικών χαρακτηριστικών και των βιομετρικών συστημάτων. Εν συνεχεία παρατίθενται τα βασικά ποιοτικά χαρακτηριστικά των βιομετρικών γνωρισμάτων, περιγράφονται αναλυτικά παραδείγματα βιομετρικών χαρακτηριστικών και παρουσιάζονται οι κυριότερες εφαρμογές τους.

3.1 *Ιστορική Ανασκόπηση Βιομετρικής Ταυτοποίησης και Επιβεβαίωσης Ταυτότητας*

Η χρήση των βιομετρικών συστημάτων δεν αποτελεί κάποια νέα μέθοδο ταυτοποίησης και επιβεβαίωσης της ταυτότητας ατόμων. Έτσι, από τις απαρχές του πολιτισμού οι άνθρωποι βασίζονταν στο ανθρώπινο πρόσωπο προκειμένου να διακρίνουν γνωστά και μη άτομα. Με τα

χρόνια η διαδικασία αυτή έγινε πιο πολύπλοκη καθώς ο πληθυσμός αυξανόταν και καθώς η εφεύρεση νέων μεθόδων μεταφοράς έφερνε νέα άτομα στις κάποτε μικρές κοινωνίες.

Στο πέρασμα των αιώνων πληθώρα ανθρώπινων χαρακτηριστικών χρησιμοποιήθηκαν ως μια πιο επίσημη μορφή ταυτοποίησης. Συγκεκριμένα, σε μια σπηλιά η οποία χρονολογείται ότι είναι παραπάνω από 31000 ετών, βρέθηκαν τοιχογραφίες οι οποίες εκτιμάται ότι δημιουργήθηκαν από προϊστορικούς άνθρωπους οι οποίοι έμεναν εκεί. Γύρω από τις τοιχογραφίες αυτές βρέθηκαν διάφορα αποτυπώματα ανθρώπινης παλάμης τα οποία θεωρείται ότι αποτελούσαν «υπογραφή» του δημιουργού τους [10]. Στοιχεία επίσης συνηγορούν ότι η χρήση των αποτυπωμάτων ως μέσο ταυτοποίησης χρονολογείται στα 500 π.Χ. όταν οι Βαβυλώνιοι αποτύπωναν τις συναλλαγές τους πάνω σε πλάκες από πηλό οι οποίες συμπεριελάμβαναν δακτυλικά αποτυπώματα [11]. Ο Ισπανός εξερευνητής και συγγραφέας Joao de Barros κατέγραψε μεταξύ των άλλων ότι οι παρόμοια χρήση των δαχτυλικών αποτυπωμάτων απαντάται και στην Κίνα του 14^{ου} αιώνα οπότε και οι έμποροι τα χρησιμοποιούσαν κατά τις εμπορικές τους συναλλαγές. Μεταξύ των άλλων βασίζονταν σε αυτό το βιομετρικό χαρακτηριστικό καθώς και το αποτύπωμα του ποδιού τα οποία ελάμβαναν μέσω μελανιού για να ξεχωρίζουν τα παιδιά [12]. Σύμφωνα με τον ίδιο συγγραφέα, οι αρχαίοι Αιγύπτιοι έμποροι βασίζονταν σε φυσικούς περιγραφείς για τον διαχωρισμό των έμπιστων εμπόρων γνωστής φήμης και με προγενέστερες επιτυχείς συναλλαγές από νέους εμπόρους.

Στα μεταγενέστερα χρόνια η χρήση βιομετρικών χαρακτηριστικών παρατηρήθηκε πιο έντονα. Το 1890 ο Alphonse Bertillon, ένας Γάλλος αστυνόμος μελέτησε τη μηχανική του σώματος καθώς και μετρήσεις προκειμένου να καταστεί δυνατή η σύλληψη εγκληματιών [13]. Η αστυνομία βασίστηκε στη μέθοδο αυτή, την επονομαζόμενη Bertillonage μέθοδο, μέχρι τη

στιγμή που αναγνώρισε λανθασμένα κάποιους ανθρώπους. Η μέθοδος αυτή γρήγορα εγκαταλείφθηκε, ενώ ξαναεισήχθει από τον Richard Edward Henry της Scotland Yard.

Στις αρχές του 20ου αιώνα ο Karl Pearson, ένας ερευνητής εφαρμοσμένων μαθηματικών, μελέτησε τη βιομετρία στο University College of London. Κατά τη διάρκεια της έρευνάς του οδηγήθηκε σε σημαντικές ανακαλύψεις στον χώρο αυτό μέσω της μελέτης στατιστικής ιστορίας και συσχέτισης, την οποία και εφάρμοσε στην εξέλιξη των ζώων. Το έργο του αυτό περιελάμβανε τη μέθοδο των στιγμών, το Pearson σύστημα των καμπύων, συσχέτιση και την chi-squared δοκιμή.

Αργότερα, κατά τις δεκαετίες 1960 και 1970, διαδικασίες βιομετρικής πιστοποίησης μέσω της υπογραφής αναπτύχθηκαν, αλλά ο βιομετρικός χώρος έμεινε σταθερός μέχρι τη στιγμή που τα στρατιωτικά γραφεία και τα γραφεία ασφαλείας προχώρησαν στην έρευνα και στην ανάπτυξη της βιομετρικής τεχνολογίας πέραν των δαχτυλικών αποτυπωμάτων.

Τα τελευταία χρόνια και με διαρκώς αυξανόμενη την προσπάθεια για βελτίωση των συστημάτων ταυτοποίησης και επιβεβαιώσης ταυτότητας τα οποία κυρίως βασίζονταν σε κωδικούς και κάρτες, συνοδευόμενα από σειρά προβλημάτων τα οποία αναλύθηκαν στο προηγούμενο κεφάλαιο, παρατηρείται μια έντονη δραστηριότητα στον χώρο των βιομετρικών συστημάτων. Μια σημαντική εφαρμογή αυτών αποτέλεσε το Super Bowl στην Tampa, Florida, το 2001 [14] κατά το οποίο έγινε χρήση συστήματος αναγνώρισης προσώπου με κάθε εικόνα προσώπου από τους 100000 θεατές να έχει καταγραφεί από κάμερες ασφαλείας και να έχει ελεγχθεί ηλεκτρονικά σε σχέση με εικόνες κακοποιών από την αστυνομία της Tampa. Αυτή η κίνηση συνδυαζόμενη από τη μη επιτυχή λειτουργία του συστήματος προκάλεσε πληθώρα αντιδράσεων από υποστηρικτές των κοινωνικών ελευθεριών οι οποίοι κατήγγειλαν τότε τις βιομετρικές τεχνολογίες ως εχθρούς της ελευθερίας του ατόμου.

Παρ'όλα αυτά και ειδικά μετά το τρομοκρατικό κτύπημα της 11ης Σεπτεμβρίου 2001 στις ΗΠΑ, πολλές αρχές παγκοσμίως προχώρησαν στην εγκατάσταση βιομετρικών τεχνολογιών στα αεροδρόμια προκειμένου να εντοπίσουν υπόπτους για τρομοκρατία. Λίγα χρόνια αργότερα και μετά από πληθώρα βομβιστικών επιθέσεων στον υπόγειο σιδηρόδρομο, η Μεγάλη Βρετανία υιοθέτησε τεχνολογίες αναγνώρισης προσώπου καταλήγοντας σήμερα το Λονδίνο να έχει περισσότερες από 200000 κάμερες σε δημόσιους χώρους για λόγους ασφαλείας.

3.2 *Ποιοτικά Χαρακτηριστικά των Βιομετρικών Γνωρισμάτων*

Όπως έχει ήδη αναφερθεί, οι βιομετρικές τεχνολογίες για ταυτοποίηση, επιβεβαίωση ταυτότητας αλλά και δημιουργία προφίλ ενός ατόμου βασίζονται σε συγκεκριμένα ανθρώπινα χαρακτηριστικά. Σε αυτό το σημείο τίθεται το ερώτημα, οι μετρήσεις ποιων ανθρώπινων χαρακτηριστικών μπορούν να χρησιμοποιηθούν για τον σκοπό αυτόν. Οποιοδήποτε ανθρώπινο φυσιολογικό ή/και συμπεριφορικό χαρακτηριστικό μπορεί να χρησιμοποιηθεί ως βιομετρικό χαρακτηριστικό αρκεί να ικανοποιεί τις ακόλουθες απαιτήσεις [1] :

- *Καθολικότητα (Universality)*: κάθε άτομο πρέπει να έχει αυτό το χαρακτηριστικό
- *Δυνατότητα Διαχωρισμού (Distinctiveness)*: κάθε δύο άτομα θα πρέπει να είναι επαρκώς διαφορετικά μεταξύ τους όσον αφορά σε αυτό το χαρακτηριστικό
- *Μονιμότητα (Permanence)*: το χαρακτηριστικό αυτό θα πρέπει να είναι επαρκώς αμετάβλητο (σε σχέση με το κριτήριο σύγκρισης) κατά τη διάρκεια μιας χρονικής περιόδου
- *Δυνατότητα συλλογής (Collectability)*: το χαρακτηριστικό αυτό θα πρέπει να μπορεί να μετρηθεί ποσοτικά

Εκτός όμως από τα παραπάνω, και λαμβάνοντας υπ'όψιν την πρακτική όψη του εγχειρήματος ανάπτυξης ενός βιομετρικού συστήματος, υπάρχουν κι άλλα θέματα που θα πρέπει να ληφθούν υπ'όψιν όπως:

- *Απόδοση (Performance)*: η οποία αφορά στη δυνατή προς επίτευξη ακρίβεια και ταχύτητα αναγνώρισης, αλλά και στους λειτουργικούς και περιβαλλοντολογικούς παράγοντες οι οποίοι επηρεάζουν την ακρίβεια και την ταχύτητα.
- *Δυνατότητα Αποδοχής (acceptability)*: η οποία δείχνει το βαθμό στον οποίο ποιοι άνθρωποι είναι πρόθυμοι να αποδεχτούν τη χρήση ενός συγκεκριμένου βιομετρικού αναγνωριστικού (identifier) στην καθημερινή τους ζωή,
- *Καταστρατήγηση (circumvention)*, η οποία αντανακλά το πόσο έυκολα μπορεί το σύστημα να ξεγελαστεί με τη χρήση δόλιων μεθόδων.

3.3 *Κατηγοριοποιήσεις Βιομετρικών Χαρακτηριστικών και Συστημάτων*

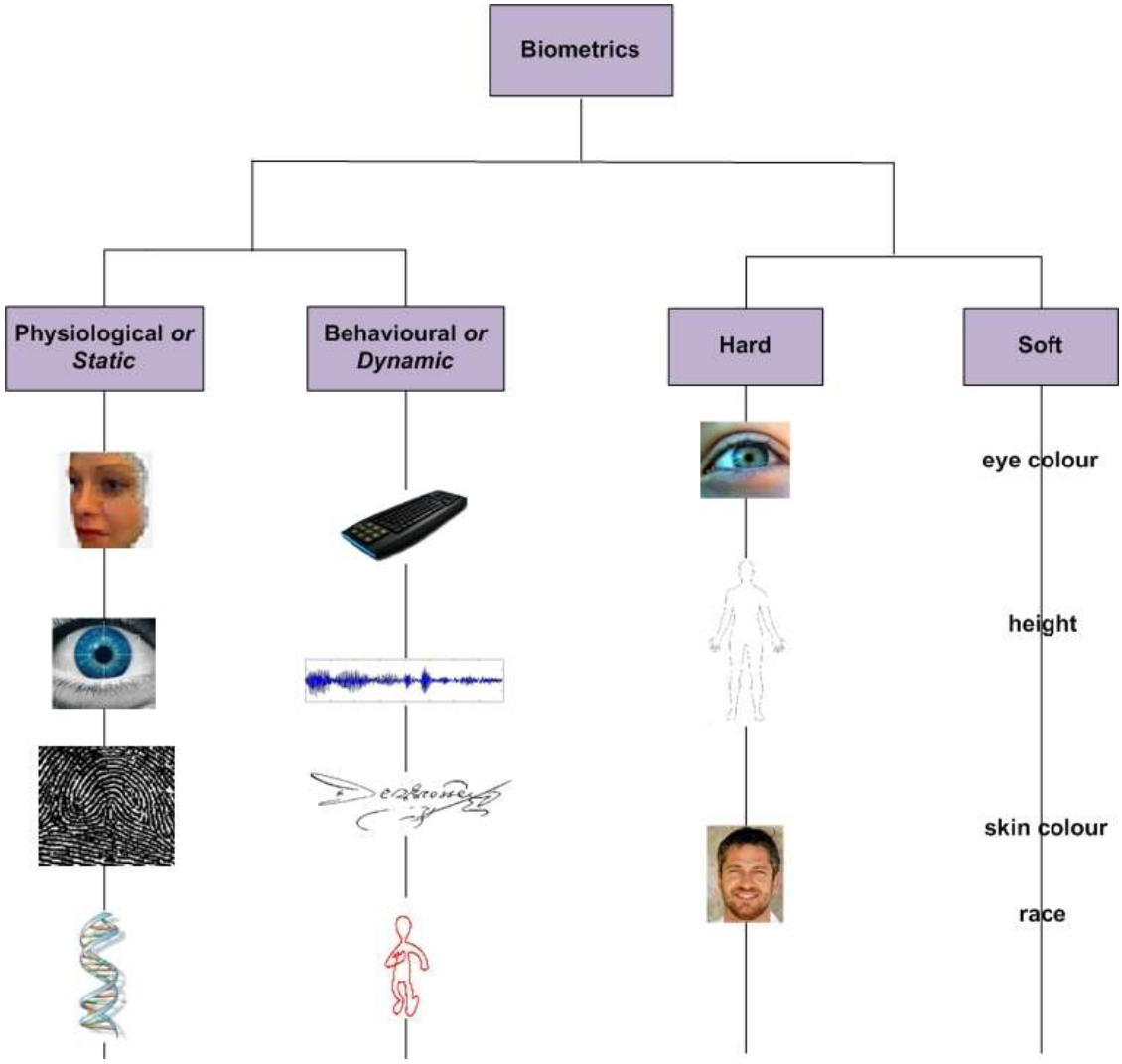
Ο πλούτος των βιομετρικών χαρακτηριστικών έχει οδηγήσει σε διάφορες κατηγοριοποιήσεις αυτών. Οι δύο κύριες κατηγορίες στις οποίες μπορούν να διακριθούν τα βιομετρικά χαρακτηριστικά είναι τα *φυσιολογικά* (ή *παθητικά*) και τα *συμπεριφορικά* (ή *ενεργητικά*). Τα φυσιολογικά βιομετρικά χαρακτηριστικά αφορούν σε ανθρώπινα χαρακτηριστικά τα οποία είναι εν γένει σταθερά ή αμετάβλητα όπως τα δαχτυλικά αποτυπώματα, η γεωμετρία του χεριού, η ίρις, η φωνή και το πρόσωπο. Τα συμπεριφορικά βιομετρικά χαρακτηριστικά σχετίζονται με δεξιότητες ή λειτουργίες που πραγματοποιούνται από ένα άτομο μια δεδομένη στιγμή και για έναν συγκεκριμένο λόγο, όπως είναι για παράδειγμα η υπογραφή ή η πληκτρολόγηση κειμένου στον υπολογιστή.

Μια ακόμη κατηγοριοποίηση των βιομετρικών χαρακτηριστικών η οποία απαντάται είναι ο διαχωρισμός τους σε *στατικά* και *δυναμικά*. Ο όρος «στατικός» αναφέρεται στη μέτρηση ενός γνωρίσματος το οποίο δεν απαιτεί κάποια δράση κατά τη στιγμή της ταυτοποίησης ή της επιβεβαίωσης ταυτότητας. Με την έννοια του «δυναμικού» προσδιορίζονται μετρήσεις ενός γνωρίσματος κατά τη διάρκεια κάποιας δραστηριότητας. Έτσι για παράδειγμα, μια υπογραφή μπορεί να μετρηθεί είτε στατικά – μέσω της εξέτασης της υπογραφής μετά την αποτύπωσή της – είτε δυναμικά – μέσω της παρατήρησης της διαδικασίας της υπογραφής.

Αναλόγως με τον βαθμό που επιτρέπουν τον διαχωρισμό μεταξύ δύο ατόμων αλλά και της μονιμότητας τους τα βιομετρικά γνωρίσματα μπορούν να διακριθούν σε *soft* και *hard biometrics*. Τα soft biometrics αφορούν σε χαρακτηριστικά τα οποία παρέχουν κάποια μορφή πληροφορίας για ένα άτομο, αλλά δεν επαρκούν για τον ουσιαστικό διαχωρισμό μεταξύ δύο ατόμων. Τα χαρακτηριστικά αυτά μπορεί να είναι είτε συνεχή (όπως είναι το ύψος και το βάρος) είτε διακριτά (όπως το φύλο, το χρώμα των ματιών, η εθνικότητα). Τα hard biometrics εμπεριέχουν όλες τις ιδιότητες που απαιτούνται για τη διάκριση μεταξύ δύο ανθρώπων. Χαρακτηριστικά παραδείγματα αυτών είναι το DNA και το δαχτυλικό αποτύπωμα. Όπως είναι προφανές, η χρήση των soft biometrics για ταυτοποίηση ή επιβεβαίωση της ταυτότητας ενός ατόμου είναι εφικτή μόνο μέσω του συνδυασμού τους με hard biometrics. Σημαντική πάντως είναι η χρησιμότητα αυτών για άλλες εφαρμογές, όπως είναι η σύνδεση ενός ατόμου με κάποιο προφίλ διατηρώντας την ανωνυμία του όπως θα δούμε και σε ακόλουθη παράγραφο.

Η κύρια κατηγοριοποίηση των βιομετρικών συστημάτων η οποία απαντάται στη βιβλιογραφία είναι ο διαχωρισμός τους σε *μονοτροπικά* (*unimodal*) και *πολυτροπικά* (*multimodal*). Στην πρώτη κατηγορία εντάσσονται τα βιομετρικά συστήματα τα οποία βασίζονται σε ένα βιομετρικό χαρακτηριστικό, όπως είναι για παράδειγμα τα συστήματα αναγνώρισης προσώπου. Αντιθέτως,

τα πολυτροπικά βιομετρικά συστήματα συνδυάζουν περισσότερα του ενός βιομετρικού χαρακτηριστικού με στόχο τη βελτίωση της ακρίβειας του συστήματος κατά την ταυτοποίηση, την επιβεβαίωση ταυτότητας ή τη σύνθεση του προφίλ του ατόμου. Παραδείγματα τέτοιων συστημάτων αποτελούν συστήματα που συνδυάζουν αναγνώριση προσώπου με αναγνώριση φωνής. Τα τελευταία χρόνια απαντάται και ο όρος διατροπικά (intramodal) ο οποίος αφορά σε βιομετρικά συστήματα που εστιάζουν σε ένα βιομετρικό χαρακτηριστικό και συμπεριλαμβάνουν fusion διαφορετικών τεχνικών σε διάφορα στάδια του βιομετρικού συστήματος.



Εικόνα 3 Κατηγορίες Βιομετρικών Χαρακτηριστικών

3.4 *Οι τρεις όψεις των Βιομετρικών Συστημάτων*

Όπως έχει ήδη αναφερθεί οι κύριες λειτουργίες ενός βιομετρικού συστήματος είναι η *ταυτοποίηση*, η *επιβεβαίωση ταυτότητας* και η *σύνθεση του προφίλ* (profiling) ενός ατόμου. Κατά τη διάρκεια της ταυτοποίησης το αποτυπωμένο βιομετρικό χαρακτηριστικό του ατόμου συγκρίνεται με όλα όσα είναι καταχωρημένα στη βάση με τα αντίστοιχα βιομετρικά δεδομένα (δεδομένα αναφοράς) προκειμένου να βρεθεί η ταυτότητα του ατόμου που προσπαθεί να αποκτήσει πρόσβαση στο σύστημα ή σε κάποια περιοχή/λειτουργία μέσω του συστήματος. Αν κατά τη διαδικασία της σύγκρισης τα εισαγμένα δεδομένα ταιριάζουν με κάποια από τα δεδομένα αναφοράς εντός ενός πλαισίου σφάλματος (το οποίο εξαρτάται από τα χαρακτηριστικά της εφαρμογής, το χρησιμοποιούμενο βιομετρικό γνώρισμα αλλά και τις απαιτήσεις του συστήματος και της εφαρμογής μεταξύ των άλλων), τότε το σύστημα επιστρέφει τα στοιχεία για το άτομο του οποίου τα δεδομένα αναφοράς είναι πιο κοντά στα βιομετρικά χαρακτηριστικά που εισήχθησαν στο σύστημα. Πρόκειται για μία 1:N (ή αλλιώς ένα προς πολλά) σύγκριση.

Κατά την επιβεβαίωση ταυτότητας τα εισηγμένα βιομετρικά χαρακτηριστικά συγκρίνονται με ένα συγκεκριμένο δείγμα βιομετρικών δεδομένων αναφοράς προκείμενου να επιβεβαιωθεί η ταυτότητα του ατόμου το οποίο προσπαθεί να αποκτήσει πρόσβαση στο σύστημα ή σε μια περιοχή/λειτουργία μέσω του συστήματος. Έτσι, όμοια με τη διαδικασία της ταυτοποίησης, αν η σύγκριση των εισηγμένων βιομετρικών δεδομένων με το δείγμα αναφοράς ταιριάζουν εντός ενός αποδεκτού πλαισίου σφάλματος, τότε επιβεβαιώνεται η ταυτότητα του ατόμου. Όπως είναι εμφανές, η διαδικασία αυτή αφορά σε μια 1:1 σύγκριση.

Όταν το βιομετρικό σύστημα ορθώς ταυτοποιεί ή επιβεβαιώνει την ταυτότητα ενός ατόμου, τότε το αποτέλεσμα της ταυτοποίησης ή της επιβεβαίωσης ταυτότητας είναι *αληθώς θετικό (true*

positive), ενώ στην περίπτωση που το σύστημα ορθώς απορρίπτει ένα άτομο ως μη ταυτιζόμενο με το άτομο το οποίο ισχυρίζεται ότι είναι, τότε πρόκειται για *αληθώς αρνητικό* (*true negative*).

Όπως έχει ήδη αναφερθεί στο προηγούμενο κεφάλαιο, η διαδικασία σύνθεσης του προφίλ έχει δύο πτυχές:

- Τη διαδικασία κατασκευής προφίλ (συσχετισμένα δεδομένα) τα οποία ταυτοποιούν και αναπαριστούν είτε ένα άτομο είτε μια ομάδα/κατηγορία
- Ή/Και την εφαρμογή των προφίλ για την αναγνώριση και αναπαράσταση ενός ατόμου ως ένα συγκεκριμένο άτομο ή ως μέλος μιας συγκεκριμένης ομάδας/κατηγορίας

Αν και εγγενώς τα βιομετρικά χαρακτηριστικά σχετίζονται με την ταυτοποίηση και την επιβεβαίωση ταυτότητας, εμπεριέχουν επίσης δεδομένα πλούσια σε πληροφορία για κατασκευή προφίλ. Καθώς η διαδικασία σύνθεσης προφίλ δεν είναι τόσο καλώς ορισμένη όσο είναι η ταυτοποίηση και η επιβεβαίωση ταυτότητας, δεν υπάρχει ακόμη σαφώς ορισμένη ερευνητική δράση για τη βιομετρική σύνθεση προφίλ. Από την οπτική της σύνθεσης προφίλ, η βιομετρική πληροφορία μπορεί είτε να χαρακτηρίζει άμεσα το υποκείμενό της, είτε να χρησιμοποιείται ως μέσο ταυτοποίησης ή επιβεβαίωσης ταυτότητας μέσω της οποίας να παρέχεται σύνδεση προς ένα υπάρχον (και πιθανότατα) μη-βιομετρικό προφίλ.

3.5 *Παραδείγματα Βιομετρικών Χαρακτηριστικών*

3.5.1 Φυσιολογικά Βιομετρικά Χαρακτηριστικά (Physiological Biometrics)

3.5.1.1 Πρόσωπο

Η αναγνώριση προσώπου αποτελεί υποπεριοχή του γενικότερου προβλήματος αναγνώρισης αντικειμένων το οποίο απαιτεί διαχωρισμό μεταξύ αντικειμένων τα οποία διαφέρουν ελάχιστα

μεταξύ τους (στην προκειμένη τα ανθρώπινα πρόσωπα) και για τον λόγο αυτόν συνιστά μεγάλη πρόκληση στον χώρο των προβλημάτων όρασης υπολογιστών. Πρόκειται για μια μη-επεμβατική βιομετρική μέθοδο η οποία βρίσκει εφαρμογή σε ποικίλους τομείς, όπως το εμπόριο (διασκέδαση, επεξεργασία ταινιών, βιντεοτηλέφωνα, τηλεδιάσκεψη), η ασφάλεια, η βιομηχανία και η επιβολή του νόμου.

Το πρώτο κύριο βήμα ενός συστήματος αναγνώρισης προσώπου είναι η ανίχνευση προσώπου στην εικόνα. Μια εικόνα μπορεί να περιέχει κανένα, ένα ή περισσότερα πρόσωπα. Έτσι, η διαδικασία ανίχνευσης και εντοπισμού του προσώπου στοχεύει στον αυτόματο και αξιόπιστο καθορισμό της θέσης όλων των περιοχών στην εικόνα οι οποίες περιέχουν ένα άτομο, ανεξαρτήτως της τρισδιάστατής του θέσης, του προσανατολισμού του, της θέσης του και των συνθηκών φωτός. Δεδομένου ότι η πολυπλοκότητα και η αποδοτικότητα της αναγνώρισης προσώπου εξαρτάται από την ακρίβεια της διαδικασίας εντοπισμού του στην εικόνα, η διαδικασία αυτή είναι μείζονος σημασίας. Έτσι, κάθε μέθοδος η οποία χρησιμοποιείται για τον εντοπισμό του προσώπου στην εικόνα θα πρέπει να λαμβάνει υπ'οψιν ότι απαντάται μεγάλη μεταβλητότητα όσον αφορά στο μέγεθος, στο χρώμα, στο σχήμα, στην υφή, στον προσανατολισμό και στην έκφραση του ανθρώπινου προσώπου.

Ο εντοπισμός προσώπου σε εικόνα αποτελεί ένα πρόβλημα για το οποίο έχει λάβει χώρα πυρετώδης ερευνητική δραστηριότητα καταλήγοντας σε πληθώρα μεθόδων εντοπισμού προσώπου οι οποίες μπορούν να κατηγοριοποιηθούν σε τέσσερεις κύριες κατηγορίες [26]:

- *Feature Invariant*: οι οποίες στοχεύουν στον εντοπισμό χαρακτηριστικών του προσώπου τα οποία υφίστανται ακόμη και όταν η οπτική γωνία, ο προσανατολισμός ή οι συνθήκες φωτισμού ποικίλουν. Τα ανιχνευμένα χαρακτηριστικά χρησιμοποιούνται για καθορισμό της θέσης του προσώπου. Βάση για τις μεθόδους αυτές αποτελεί η παρατήρηση ότι οι

άνθρωποι έχουν τη δυνατότητα ανίχνευσης προσώπων υπό ποικίλες συνθήκες και διαφορετικές περιπτώσεις προσανατολισμού, κι έτσι πρέπει να υπάρχουν χαρακτηριστικά ή οι ιδιότητες που δεν επηρεάζονται από αυτές τις μεταβαλλόμενες συνθήκες και καταστάσεις. Παραδείγματα τέτοιων χαρακτηριστικών αποτελούν η υφή του προσώπου (συμπεριλαμβάνοντας τις τρίχες και το δέρμα) και το χρώμα του προσώπου. Το τελευταίο χρησιμοποιείται έχοντας ως βάση μελέτες οι οποίες δείχνουν ότι η διαφορά στο χρώμα του προσώπου οφείλεται κυρίως στη διαφορά της έντασης (intensity) παρά στη διαφορά του χρώματος (chrominance) καλήγοντας σε πληθώρα μοντέλων του χρώματος του ανθρώπινου προσώπου τα οποία έχουν αναπτυχθεί σε διάφορους χρωματικούς χώρους (colour spaces). Τα τελευταία χρόνια πλήθος τεχνικών οι οποίες συνδυάζουν χαρακτηριστικά του προσώπου προκειμένουν να εντοπίσουν πρόσωπα έχουν παρουσιαστεί. Τα χαρακτηριστικά αυτά περιλαμβάνουν το χρώμα του δέρματος, το σχήμα και το μέγεθος για τον καθορισμό των υποψήφιων περιοχών στην εικόνα, ενώ ακολουθεί η φάση της επιβεβαίωσης στις υποψήφιες περιοχές κατά τις οποίες εξετάζεται η ύπαρξη χαρακτηριστικών του προσώπου όπως είναι η μύτη, το στόμα, τα φρύδια και τα μαλλία.

- *Μέθοδοι βασισμένες σε Γνώση (Knowledge-based)*: οι οποίες βασίζονται στην προσπάθεια κωδικοποίησης της ανθρώπινης γνώσης όσον αφορά στο πώς μοιάζει ένα συνηθισμένο ανθρώπινο πρόσωπο. Τις περισσότερες φορές αυτή η γνώση αποτελείται από κανόνες οι οποίοι αποτυπώνουν τις σχέσεις μεταξύ των χαρακτηριστικών του προσώπου σχετικά με τις σχετικές τους αποστάσεις και θέσεις. Παραδείγματα τέτοιων κανόνων είναι το γεγονός ότι ένα πρόσωπο έχει δύο μάτια τα οποία είναι συμμετρικά το ένα ως προς το άλλο, μια μύτη και ένα στόμα. Όμως, εμφωλεύει ο κίνδυνος αν οι

κανόνες αυτοί είναι πολύ αυστηροί να μην καταφέρουν να ανιχνεύσουν ένα πρόσωπο το οποίο δεν τους ικανοποιεί, ενώ αν είναι πολύ γενικοί, μπορεί να παρέχουν πολλά λανθασμένα θετικά. Επιπροσθέτως, ο εντοπισμός προσώπων σε διαφορικές στάσεις μεγαλώνει σημαντικά το σύνολο των περιπτώσεων τις οποίες οι κανόνες θα πρέπει να ενσωματώνουν. Παρ'όλα αυτά, στις περιπτώσεις των εικόνων που περιέχουν την εμπρόσθια όψη προσώπων, οι μέθοδοι αυτές μπορούν να εντοπίσουν αρκετά αποδοτικά τα ανθρώπινα πρόσωπα.

- *Μέθοδοι βασισμένες στην εμφάνιση (Appearance-based methods)*: οι οποίες βασίζονται σε μοντέλα τα οποία προκύπτουν από ένα μεγάλο σύνολο εικόνων προσώπου ποικίλων χαρακτηριστικών που χρησιμοποιούνται για εκπαίδευση (training) με τα χαρακτηριστικά του προσώπου τα οποία προκύπτουν από τη διαδικασία μάθησης να αποθηκεύονται υπό τη μορφή είτε μοντέλων κατανομής είτε συναρτήσεων διάκρισης. Τα μοντέλα αυτά χρησιμοποιούνται για εντοπισμό προσώπου με την εφαρμογή τους στις υπό εξέταση εικόνες.
- *Μέθοδοι σύγκρισης προτύπων*: οι οποίες περιλαμβάνουν τη σύγκριση μιας εικόνας εισόδου με δεδομένα πρότυπα ενός προσώπου τα οποία περιγράφουν είτε το πρόσωπο ολόκληρο είτε τα χαρακτηριστικά του προσώπου ξεχωριστά. Κατά τη διαδικασία αυτήν λαμβάνει χώρα ο υπολογισμός των τιμών συσχέτισης και έτσι η παρουσία ενός προσώπου στην εικόνα εισόδου καθορίζεται από τις τιμές αυτές.

Η αναγνώριση προσώπου λαμβάνει υπ'όψιν το αποτέλεσμα του εντοπισμού προσώπου και στοχεύει στην ταυτοποίηση ή στην επιβεβαίωση της ταυτότητας του ατόμου που έχει εντοπιστεί στην εικόνα. Οι υπάρχουσες μέθοδοι αναγνώρισης προσώπου μπορούν να διακριθούν σε δύο μεγάλες κατηγορίες : τις αναλυτικές και τις ολιστικές μεθόδους [25] [28]. Οι αναλυτικές ή

βασισμένες σε χαρακτηριστικά μέθοδοι (feature-based) εστιάζουν στη μελέτη της εξαγωγής χαρακτηριστικών στον χώρο και υπολογίζουν ένα σύνολο από γεωμετρικά χαρακτηριστικά από το πρόσωπο όπως τα μάτια, η μύτη και το στόμα. Οι ολιστικές ή βασισμένες στην εμφάνιση (appearance-based) μέθοδοι λαμβάνουν υπ'όψιν τις γενικές ιδιότητες του ανθρώπινου προσώπου. Το πρόσωπο αναγνωρίζεται συνολικά χωρίς τη χρήση μόνο συγκεκριμένων σημείων αναφοράς τα οποία λαμβάνονται από διαφορετικές περιοχές του προσώπου. Οι μέθοδοι αυτές γενικώς λειτουργούν απευθείας στην αναπαράσταση σε πίνακα της έντασης των εικονοστοιχείων (pixel intensity) χωρίς τον εντοπισμό των χαρακτηριστικών του προσώπου. Δεδομένου ότι δεν απαιτείται ανίχνευση γεωμετρικών χαρακτηριστικών του προσώπου, το σύνολο των μεθόδων αυτών είναι συνήθως πιο πρακτικό και πιο εύκολο για υλοποίηση σε σχέση με γεωμετρικές βασισμένες σε χαρακτηριστικά μεθόδους. Ερευνητικές προσπάθειες για συνδυασμό αναλυτικών και ολιστικών μεθόδων επίσης έχουν λάβει χώρα [27].

3.5.1.2 Δαχτυλικά Αποτυπώματα

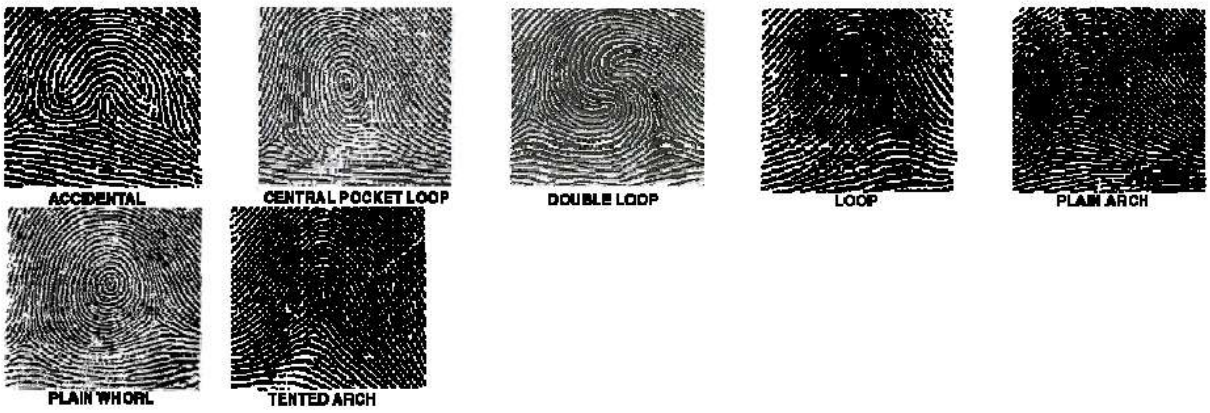
Όπως έχει ήδη αναφερθεί, η ιδέα ότι η χρήση δακτυλικών αποτυπωμάτων μπορεί να οδηγήσει σε ταυτοποίηση ή επιβεβαίωση της ταυτότητας ενός ατόμου αποτελεί μια μέθοδο η οποία απαντάται ανά τους αιώνες. Η τεχνολογική έκρηξη που παρατηρήθηκε τον $20^{ο}$ αιώνα έδωσε πρόσφορο έδαφος για την ανάπτυξη αυτομάτων συστημάτων αναγνώρισης δακτυλικών αποτυπωμάτων τα οποία δοκιμάστηκαν και εισήχθησαν αρχικά στις Ηνωμένες Πολιτείες Αμερικής και στην Αυστραλία. Η αρχική αυτή εισαγωγή τέτοιων βιομετρικών συστημάτων άνοιξε τον δρόμο για τη γενικότερη χρήση των δακτυλικών αποτυπωμάτων ως μέσο ταυτοποίησης και πιστοποίησης της ταυτότητας η οποία απαντάται σε πληθώρα χώρων, όπως επίσημη ταυτοποίηση ατόμων από τις αστυνομικές αρχές και ο έλεγχος πρόσβασης σε υπολογιστές μεταξύ των άλλων.

Τα δακτυλικά αποτυπώματα δεν καθορίζονται γενετικά, αλλά ξεκινούν να αναπτύσσονται τυχαία κατά τον 3^{ο} μήνα της κύησης στις άκρες των δακτύλων των εμβρύων [16]. Η τεχνική αναγνώρισης δακτυλικών αποτυπωμάτων βασίζεται σε αυτές τις ποικίλες θηλοειδείς δομές στα άκρα των ανθρώπινων δακτύλων. Η μοναδικότητα των δακτυλικών αποτυπωμάτων μεταξύ των ατόμων βασίζεται στην ποικιλία διακριτών χαρακτηριστικών, όπως είναι οι πόροι και οι ραβδώσεις. Έτσι, για παράδειγμα, το πλήθος των πόρων, τα σχετικά τους σχήματα και μεγέθη, οι σχετιές τους θέσεις αποτελούν παράγοντες που ενισχύουν τη μοναδικότητα αυτών των χαρακτηριστικών.

Ρίχνωντας μια πιο κοντινή ματιά στη μορφολογία των άκρων των δακτύλων, οι πόροι σχηματίζονται όταν οι ιδρωτοποιοί αδένες στο υποδόριο στρώμα του δέρματος παράγουν πόρους ιδρώτα οι οποίοι με τη σειρά τους μεγαλώνουν διαμέσου του υποδόριου στρώματος και του δέρματος στην επιδερμίδα [19]. Οι επιδερμικές ραβδώσεις ξεκινούν να σχηματίζονται κατά τον 6^{ο} μήνα της κύησης, οπότε και το πρόπυπο το οποίο σχηματιζόταν στην περιοχή της αδενικής πτυχής μεταφέρεται στην επιδερμίδα. Πολλοί παράγοντες επηρεάζουν τον σχηματισμό του πρότυπου της επιδερμίδας [20], όπως είναι η σταθεροποίηση η οποία λαμβάνει χώρα όταν οι πόροι έκκρισης των ιδρωτοποιών αδένων ανοίγουν στην επιφάνεια, κατά τακτά χρονικά διαστήματα, στις θηλακοειδείς ραβδώσεις. Αυτά τα ανοίγματα στην επιφάνεια είναι πόροι και η συχνότητα εμφάνισης αυτών αποτελεί σημαντική παράμετρο που καθορίζει τη μοναδικότητα των συνθέσεων των πόρων. Η θέση των πόρων είναι σταθερή μόλις ολοκληρωθεί ο σχηματισμός τους στη ράβδωση. Έρευνες οι οποίες έχουν πραγματοποιηθεί έχουν καταλήξει ότι οι πόροι δεν εξαφανίζονται, μετακινούνται ή δημιουργούνται με το πέρασμα του χρόνου [21].

Αναλόγως το σύστημα, επτά με εννέα τυπικά πρότυπα μπορούν να διακριθούν [17], τα οποία μπορούν να κατηγοριοποιηθούν γενικώς σε τόξο (arch), έλικα (whorl) και βρόχο (loop), όπως

παρουσιάζονται και στην Εικόνα 4. Πέραν των κύριων χαρακτηριστικών των δακτυλικών αποτυπωμάτων χρησιμοποιούνται και δευτερεύοντα χαρακτηριστικά όπως είναι η θέση των απολήξεων των ραβδώσεων καθώς και οι διακλαδώσεις τους.



Εικόνα 4 Επτά κοινά πρότυπα που χρησιμοποιούνται από το FBI [18]

Η λήψη των δακτυλικών αποτυπωμάτων μπορεί να γίνει με διαφόρων ειδών αισθητήρες, όπως όπως οπτικούς αισθητήρες, αισθητήρες ηλεκτρομαγνητικού πεδίου, θερμικούς αισθητήρες και υπερηχητικούς αισθητήρες. Όπως θα δούμε και στην Τα τρωτά σημεία των βιομετρικών συστημάτων, νέοι εξελιγμένοι αισθητήρες οι οποίοι ανιχνεύουν κατά τη διάρκεια της λήψης των δακτυλικών αποτυπωμάτων αν τα δεδομένα λαμβάνονται από ζώντα οργανισμό (liveness detection) έχουν αναπτυχθεί με στόχο την αποφυγή περιπτώσεων «εξαπάτησης» των συστημάτων αναγνώρισης δακτυλικών αποτυπωμάτων με την παρουσίαση ψεύτικων ή κλεμμένων δακτυλικών αποτυπωμάτων.

3.5.1.3 DNA

Η γενετική αναγνώριση, η οποία είναι γνωστή και ως αναγνώριση βασισμένη στο DNA, αποτελεί μια μέθοδο ταυτοποίησης η οποία αναπτύχθηκε τα τελευταία χρόνια. Η μέθοδος αυτή βασίζεται στη μοναδικότητα του DNA μεταξύ των ανθρώπων, εκτός από τις περιπτώσεις ομοζυγωτών διδύμων. Έτσι, αν και όλοι οι άνθρωποι μοιραζόμαστε το 99,9% του γονιδιώματός

μας, το 0,1% των 3 δισεκατομμυρίων νουκλεοτιδίων αποτελεί ένα σημαντικό και ανιχνεύσιμο επίπεδο διαφοράς. Το μεγαλύτερο μέρος αυτής της διαφοροποίησης εντοπίζεται στις περιοχές μη-κωδικοποίησης του DNA (τις λεγόμενες «σκάρτες» περιοχές), με τη μεγάλη πρόκληση να έγκειται στην εύρεση αυτών των διαφορών.

Η ανάλυση του DNA αποτελεί μια εξαιρετικά χρήσιμη μέθοδο για πολλούς λόγους. Εν πρώτοις το DNA περιέχει «ανιχνεύσιμα» τμήματα τα οποία είναι μοναδικά σε κάθε άτομο. Συγκεκριμένα, η πιθανότητα δύο άτομα να έχουν ακριβώς το ίδιο προφίλ DNA έχει τιμή μεταξύ 1 στα 5000000 και 1 στα 100 δισεκατομμύρια, δηλαδή μεταγαλύτερη από τον ανθρώπινο πληθυσμό στη γη, με την τιμή αυτής να μεγαλώνει ακόμη περισσότερο με το πλήθος των περιοχών του DNA οι οποίες εξετάζονται. Επιπροσθέτως, το μόριο του DNA είναι αρκετά εύρωστο και σταθερό υπό ποικίλες (αν και όχι όλες) περιβαλλοντολογικές συνθήκες, ενώ είναι δυνατή η απομόνωσή του από ένα μεγάλο εύρος βιολογικών δειγμάτων. Κάθε κύτταρο ενός ανθρώπου περιέχει το ίδιο DNA με αποτέλεσμα η πηγή λήψης του να μην έχει σημασία, ενώ ακόμη και με εξαιρετικά μικρές ποσότητες δείγματος (όπως μια σταγόνα αίματος, μια τρίχα ή το σάλιο πάνω σε ένα τσιγάρο) είναι δυνατή η ανίχνευση των τμημάτων που απαιτούνται για την ανάλυσή του. Παράλληλα, οι μέθοδοι που έχουν αναπτυχθεί για την ανάλυσή του είναι γρήγορες και σχετικά οικονομικές, ενώ τα δεδομένα της ανάλυσης μπορούν εύκολα να αποθηκευτούν σε βάσεις δεδομένων και να ανακτηθούν αποδοτικά. Συνήθεις πηγές από τις οποίες λαμβάνεται δείγμα προς ανάλυση DNA είναι το αίμα, το σπέρμα, οι τρίχες, το δέρμα, σταγόνες ιδρώτα και οι ρινικές εκροές.

Για την ανάλυση DNA τέσσερα μιτοχονδριακά μόρια DNA και τα μόρια DNA τα οποία είναι αποθηκευμένα στα χρωμοσώματα στον πυρήνα των κυττάρων χρησιμοποιούνται. Στον χώρο τα μεγάλα μόρια του DNA έχουν τη μορφή δύο επιμήκων αλυσίδων οι οποίες συστρέφονται

ελικοειδώς μεταξύ τους. Το κάθε μόριο DNA περιέχει τέσσερις χημικές μονάδες (πρωτεινικές βάσεις): την αδενίνη (A), τη γουανίνη (G), την κυτοσίνη (C) και τη θυμίνη (T). Το μόριο του DNA μοιάζει με μία σκάλα: τα πλαϊνά τμήματα αποτελούνται από ένα σκελετό σακχάρου και φωσφόρου και τα σκαλιά δημιουργούνται από το δέσμο ανάμεσα στις συμπληρωματικές βάσεις (A με T και G με C). Η θέση και η διαδοχή των βάσεων σε αυτή τη δομή κωδικοποιούν τη γενετική πληροφορία.

Η ανάλυση DNA προς ταυτοποίηση βασίζεται στα μη-κωδικοποιημένα τμήματα του DNA, δηλαδή τα τμήματα που δεν περιέχουν γενετική πληροφορία. Οι επονομαζόμενες δυαδικές επαναλήψεις (tandem repeats), οι οποίες αποτελούνται από επαναλήψεις της ίδιας αλληλουχίας βάσεων, ποικίλουν στον αριθμό των επαναλήψεων σε μεγάλο βαθμό μεταξύ διαφορετικών ανθρώπων, με εξαίρεση την περίπτωση των ομοζυγωτών διδύμων. Η θέση των επαναλήψεων στο γονιδίωμα καλείται locus (πληθ. loci). Οι κυρίαρχες δυαδικές επαναλήψεις οι οποίες χρησιμοποιούνται για την ανάλυση DNA για ταυτοποίηση είναι δύο:

a) α) Οι δυαδικές επαναλήψεις, γνωστές και ως μεταβλητού αριθμού δυαδικές επαναλήψεις (Variable Number of Tandem Repeats - VNTR) ποικίλουν από βασικές μονάδες των 2 ως και 103 βάσεων και 9 ως 100 επαναλήψεων
b) β) Οι Μικρές δυαδικές επαναλήψεις (Short Tandem Repeats - STR), οι οποίες ποικίλουν από 1 ως 30 επαναλήψεις. Από τα 105 γνωστά loci των STR περίπου 20 χρησιμοποιούνται για γενετική ταυτοποίηση. Οι STR απαντώνται ευρέως στο γονιδίωμα και μερικές από αυτές γειτνιάζουν με κωδικοποιημένα τμήματα του DNA[24].

Μια ιδιαιτέρως διαδεδομένη και σχετικά απλή μέθοδος ανάλυσης DNA είναι η ηλεκτροφόρεσις. Η μέθοδος αυτή βασίζεται στο γεγονός ότι τα μόρια του DNA περιέχουν πολλά υπόλοιπα φωσφορικού άλατος, με αποτέλεσμα να είναι αρνητικώς φορτισμένα εντός αλκαλικού

διαλύματος. Έτσι, όταν βρίσκονται σε ηλεκτρικό πεδίο, κινούνται προς το θετικό ηλεκτρόδιο, την άνοδο. Αν αυτή η μετακίνηση λάβει χώρα μέσα σε ένα τζελ, τότε τα DNA μόρια παράλληλα διαχωρίζονται με βάση το μέγεθός τους, αφού τα μικρότερα μόρια θα κινούνται με μεγαλύτερη ταχύτητα, με το τζελ έτσι να δρα ως μια «σίτα» μορίων [22]. Όπως κάθε βιοχημική ή χημική ανάλυση, η ποιότητα των αποτελεσμάτων εξαρτάται σημαντικά από την ποιότητα των δειγμάτων τα οποία αναλύονται (όπως είναι αναμεμειγμένα δείγματα DNA από διαφορετικούς ανθρώπους) καθώς και από τις διαδικασίες οι οποίες ακολουθούνται κατά τη διάρκεια της ανάλυσης [23].

Τα τελευταία χρόνια η μέθοδος ανάλυσης του DNA έχει βρει εφαρμογή σε πληθώρα τομέων. Όσον αφορά στους ανθρώπους, τα πιο δημοφιλή παραδείγματα εφαρμογής της είναι τα τεστ πατρότητας και οι εγκληματολογικές εξετάσεις [23]. Όσον αφορά στις εγκληματολογικές αναλύσεις, αναλόγως με το αναλυόμενο δείγμα και το περιβάλλον, οι μέθοδοι αυτές στις περισσότερες περιπτώσεις παρέχουν μόνο ένδειξη ότι το ταυτοποιημένο άτομο ήταν στη σκηνή εγκλήματος. Μια θετική αναγνώριση δε μπορεί να τις περισσότερες φορές να αποδείξει ότι το άτομο που αναγνωρίστηκε μέσω της γενετικής μεθόδου ότι όντως διέπραξε το έγκλημα, καθώς όπως και στις υπόλοιπες βιομετρικές μεθόδους υπάρχουν διάφοροι τρόποι «εξαπάτησης», όπως για παράδειγμα με την τοποθέτηση τριχών στον τόπο του εγκλήματος από άτομο το οποίο δεν ήταν παρόν.

3.5.2 Συμπεριφορικά Βιομετρικά Χαρακτηριστικά(Behavioural Biometrics)

3.5.2.1 Περπάτημα

Το ανθρώπινο περπάτημα, ως αποτέλεσμα πολλών συγχρονισμένων κινήσεων ενός μεγάλου αιρθμού ανθρώπινων συνδέσμων και μυών, ενσωματώνει ποικίλα χαρακτηριστικά και ιδιότητες

– είτε στατικά είτε μεταβλητά – τα οποία εξαρτώνται από διάφορους ψυχολογικούς και φυσιολογικούς παράγοντες οι οποίοι διαφέρουν μεταξύ των ανθρώπων καθώς και από εξωτερικές επιρροές όπως είναι το έδαφος, η υγεία του ατόμου, τα ρούχα και η διάθεση [29]. Η ανάλυση του ανθρώπινου περπατήματος (gait analysis) έχει δείξει ότι ο τρόπος που άνθρωπος περπατάει ενθηλακώνει ικανότητες αναγνώρισης, κι έτσι, αποτελεί ένα σχετικά νέο αλλά πολλά υποσχόμενο συμπεριφορικό βιομετρικό χαρακτηριστικό. Η μη επεμβατική του φύση – βασισμένη στη δυνατότητα συλλογής αυτού χωρίς τη συναίνεση του υποκειμένου – το έχει αναδείξει σε μια πολλά υποσχόμενη εναλλακτική μέθοδο για παθητική παρακολούθηση, ενώ, καθώς εξαρτάται από τα ανθρώπινα χαρακτηριστικά όπως το ύψος, το φύλο και τη δομή του σώματος, ενσωματώνει χρήσιμη πληροφορία τόσο για ταυτοποίηση όσο κυρίως για σύνθεση του προφίλ ενός ατόμου. Οι χαμηλές της απαιτήσεις σε συσκευές αισθητήρων και περιβάλλουσες συνθήκες – οι μετρήσεις μπορούν να λάβουν χώρα ακόμη και σε εικόνες χαμηλής ανάλυσης ή μερικής κάλυψης (partial occlusion) – αποτελούν σημαντικά της πλεονεκτήματα [40].

Η πρόοδος που έχει σημειωθεί σε αυτό το ερευνητικό πεδίο είναι στενά συνδεδεμένο με την έρευνα στο πεδίο εντοπισμού ατόμου, ανίχνευσης ανθρώπινης κίνησης, παρακολούθηση ατόμου και αναγνώριση δραστηριότητας. Όπως και τα περισσότερα προβλήματα αναγνώρισης, η αναγνώριση του ανθρώπινου περπατήματος περιλαμβάνει δύο φάσεις: τη *φάση εξαγωγής χαρακτηριστικών* (*feature extraction phase*) κατά την οποία η πληροφορία κίνησης εξάγεται από τα ληφθέντα δεδομένα και περιγράφεται με ορισμένη μορφή και τη *φάση σύγκρισης* (*matching phase*) η οποία περιλαμβάνει την κατηγοριοποίηση των προτύπων κίνησης τα οποία προέκυψαν από την πρώτη φάση.

Τα εξαγμένα χαρακτηριστικά μπορεί είτε να είναι βασισμένα στην εμφάνιση είτε περιγραφείς κίνησης (motion descriptors). Στην πρώτη προσέγγιση είτε ολόκληρο το περίγραμμα του

ανθρώπινου σώματος χρησιμοποιείται [30] [31] [35] [36] είτε τα δεδομένα αυτά [32] [33] [34] μειώνονται μέσω για παράδειγμα της προβολής δεδομένων, ενώ στη δεύτερη προσέγγιση οι περιγραφείς κίνησεις αποτελούν τον συνδυασμό των παραμέτρων όπως μέγεθος διασκελισμού και γωνίες άκρων, καθώς και παραμέτρων σώματος οι οποίες καθορίζουν την κίνηση (ύψος, μήκος άκρων, κλπ). Οι παράμετροι αυτές θα μπορούσαν να θεωρηθούν και παράμετροι του προφίλ του ατόμου. Κατά τη φάση σύγκρισης χρησιμοποιείται μια μέθοδος κατηγοριοποίησης προτύπων, όπως τα Κρυμμένα Μαρκοβιανά Μοντέλα (Hidden Markov Models-HMM) [37] [38] [39]. Παρ'όλα αυτά οι περισσότερες από αυτές τις προσεγγίσεις έχουν κυρίως δοκιμαστεί σε μικρές και/ή όχι ρεαλιστικές βάσεις δεδομένων και αντιμετωπίζουν περιορισμούς εξαιτίας θορύβου, απουσίας χωρικών δεδομένων, ακρίβειας στις εξαγμένες παραμέτρους κίνησης, κλπ.

3.5.2.2 Υπογραφή

Οι χειρόγραφες υπογραφές αποτελούν ένα ευρέως αποδεκτό μέσον για επιβεβαίωση της ταυτότητας των ατόμων για κρατικές λειτουργίες, έλεγχο αυθεντικότητας εγγράφων και οικονομικές συναλλαγές μεταξύ των άλλων. Οι άνθρωποι είναι σε θέση να αναγνωρίζουν την υπογραφή τους με μια ματιά, ενώ η επιβεβαίωση της υπογραφής συνήθως λαμβάνει χώρα υπό εποπτεία. Όμως, εξαιτίας των απαιτήσεων σε χρόνο και προσπάθεια, η επιβεβαίωση υπογραφής συχνά παρακάμπτεται οδηγώντας σε περιπτώσεις μη ανιχνευμένης πλαστογραφίας, η οποία μπορεί να αφορά σε επίσημες συμφωνίες, συμβόλαια, κλπ. Για τον λόγο αυτόν, τα τελευταία χρόνια η ανάγκη για αυτόματη επιβεβαίωση υπογραφής λαμβάνει διαρκώς αυξανόμενες διαστάσεις. Ένα τέτοιο σύστημα είναι ικανό να επιβεβαιώσει την ταυτότητα του ατόμου τόσο εξετάζωντας την υπογραφή του ατόμου όσο και τον τρόπο με τον οποίο το άτομο τη σημειώνει και συγκρίνοντάς τα δεδομένα αυτά με τις καταχωρημένες υπογραφές [41].

Τα αυτόματα συστήματα επιβεβαίωσης υπογραφής μπορούν να διακριθούν σε off-line συστήματα τα οποία εξετάζουν μόνο προγενέστερες χειρόγραφες υπογραφές και on-line συστήματα τα οποία συλλέγουν και επεξεργάζονται επιπρόσθετη πληροφορία κατά τη διάρκεια της σημείωσης της υπογραφής, όπως τη χρονική διάρκεια που απαιτείται για να σημειώσει το άτομο την υπογραφή του, την ταχύτητα και την επιτάχυνση των κινήσεων του χεριού, μέσω κατάλληλων συσκευών. Τα off-line συστήματα συλλέγουν λιγότερη πληροφορία από τα on-line, ενώ η δυναμική πλευρά της διαδικασίας σημείωσης της υπογραφής αποτελεί μια πολύ πιο δύσκολη διαδικασία [44] – αν είναι εφικτή – και είναι σημαντικά κατώτερης ποιότητας.

Στην πρώτη περίπτωση, ένα σύνολο χαρακτηριστικών τα οποία θεωρούνται σταθερά εξάγονται από την εικόνα της υπογραφής [42] [43] [45], όπως το πλήθος των κλειστών γραμμών, η αναλογία του ύψους των ψηλών γραμμάτων προς τα μικρά γράμματα και η απόσταση μεταξύ των γραμμάτων. Στη δεύτερη περίπτωση συστημάτων επιβεβαίωσης υπογραφής απαιτείται ειδικός εξοπλισμός (ο οποίος για παράδειγμα να μετράει πίεση) προκειμένου να εξαχθούν δυναμικές ιδιότητες μιας υπογραφής επιπροσθέτως του σχήματός της [46] [47] [48]. Οι ιδιότητες αυτές περιλαμβάνουν τον χρόνο που απαιτείται από το άτομο για να υπογράψει, την πίεση που ασκείται από τον στυλογράφο, το πλήθος των φορών που ο στυλογράφος ανασηκώνεται από το χαρτί, οι κατευθύνσεις κατά τη γραφή, και άλλες. Ανεξαρτήτως αν το σύστημα είναι on-line ή off-line, τα εξαγόμενα χαρακτηριστικά αποδίδονται σε μία από τις κατηγορίες: αυθεντικό ή πλαστό. Οι πιο γνωστοί ταξινομητές (classifiers) οι οποίοι χρησιμοποιούνται είναι οι Support Vector Machines (SVM) [49] [52] και τα HMM [50] [51].

Τα συστήματα επιβεβαίωσης υπογραφής, πέραν των περιορισμών που προκύπτουν από την πραγματική επίδοση των τρεχουσών μεθόδων αναγνώρισης προτύπων και εξαγωγής

χαρακτηριστικών, αντιμετωπίζουν προβλήματα τα οποία προέρχονται από τις παραλλαγές στον τρόπο με τον οποίο ένα άτομο μπορεί να υπογράψει με το πέρασμα του χρόνου. Για τον λόγο αυτόν, απαιτούνται περιοδικές ενημερώσεις των καταχωρημένων υπογραφών. Επιπροσθέτως, προκειμένου να διεξάγονται αξιόπιστοι έλεγχοι απάτης απαιτούνται αξιόπιστα δεδομένα εισόδου. Έτσι, πλαστογραφημένες υπογραφές από ειδικούς θα προσέφεραν μεγάλη αξία στα συστήματα αυτά, όμως τις περισσότερες φορές τέτοια δεδομένα είναι δυσεύρετα.

3.6 *Βιομετρικές εφαρμογές*

Οι βιομετρικές εφαρμογές μπορούν γενικώς να κατηγοριοποιηθούν σε:

1. ***Κυβερνητικού ελέγχου μοντέλα ταυτοποίησης***: για τις εφαρμογές αυτές μια δημόσια αρχή αναλαμβάνει να συλλέξει βιομετρικά δεδομένα προκειμένου να τα συμπεριλάβει σε μια εφαρμογή ταυτοποίησης, όπως είναι οι ταυτότητες, τα διαβατήρια και οι κάρτες κοινωνικής ασφάλισης. Ο έλεγχος επί των δεδομένων αυτών μπορεί να είναι κεντρικός, να διαιρείται σε περισσότερους του ενός οργανισμούς αλλά υπό κατάλληλες συμφωνίες ή πολυμερής χωρίς τις απαραίτητες συμφωνίες σχετικά με την αποκάλυψη ή τη μεταφορά των βιομετρικών δεδομένων.
2. ***Μοντέλο ελέγχου πρόσβασης***: σε αυτήν την κατηγορία μια δημόσια ή ιδιωτική αρχή αναλαμβάνει τη συλλογή των βιομετρικών δεδομένων προκειμένου να διασφαλίσει την πρόσβαση σε μια τοποθεσία ή σε μια διαδικτυακή εφαρμογή. Ο έλεγχος επί των δεδομένων μπορεί να είναι κεντρικός ή διαιρεμένος σε περισσότερους του ενός οργανισμούς αλλά με κατάλληλες συμφωνίες ή να διαιρείται έτσι ώστε το άτομο να μοιράζεται τον έλεγχο.

3. ***Μικτό μοντέλο*** : στο μοντέλο αυτό τα βιομετρικά δεδομένα που έχουν συλλεγεί θα μοιράζονται και να ανταλλάσσονται μετακύ δημόσιων και ιδιωτικών αρχών.
4. ***Μοντέλο διευκόλυνσης*** : είτε το υποκείμενο αναλαμβάνει εξ ολοκλήρου την απόφαση χρήσης των βιομετρικών δεδομένων για αποκλειστικά ιδιωτικούς σκοπούς διευκόλυνσης, είτε ένας οργανισμός χρησιμοποιεί τα βιομετρικά δεδομένα προς απλούστευση της διαχειριστικής διαδικασίας είτε με κεντρικό είτε με διαμοιρασμένο έλεγχο.
5. ***Μοντέλο Παρακολούθησης*** : μια δημόσια ή ιδιωτική αρχή αναλαμβάνει να συλλέξει και να επεξεργαστεί βιομετρικά δεδομένα για σκοπούς παρακολούθησης.

Στον ακόλουθο πίνακα παρατίθενται ενδεικτικές εφαρμογές των βιομετρικών συστημάτων σε διάφορους χώρους:

Ρυθμίζουσα αρχή	**Στόχος**	**Ενδεικτικές Εφαρμογές**
Δημόσια αρχή (εθνική ή τοπική κυβέρνηση)	Έλεγχος πρόσβασης	Έλεγχος ταξιδιωτών Έλεγχος πρόσβασης σε κυβερνητικές εγκαταστάσεις
Ιδιωτικός Οργανισμός/Τράπεζες		Έλεγχος πρόσβασης σε κτηριακές εγκαταστάσεις Έλεγχος πρόσβασης σε εξοπλισμό και πληροφοριακά συστήματα
Εμπορική Εταιρεία		Έλεγχος πρόσβασης σε καταστήματα/μαγαζιά διασκέδασης/θεματικά πάρκα
Ιδιώτης		Έλεγχος πρόσβασης σε ιδιοκτησία (υπολογιστές, αυτοκίνητα, οικίες)
Δημόσια αρχή (εθνική ή τοπική κυβέρνηση)	Ασφάλεια	Επιβεβαίωση ταυτότητας εγκληματιών Εντοπισμός υπόπτων Εξιχνίαση εγκλημάτων
Ιδιωτικός Οργανισμός		Εντοπισμός ύποπτης δραστηριότητας
Ιδιώτης	Παρακολούθηση κατάστασης υγείας/φυσικής κατάστασης	Αυτόματη μέτρηση επιπέδων κούρασης οδηγών Αναγνώριση συναισθηματικής κατάστασης αρρώστων/ηλικιωμένων

Πίνακας 1 Ενδεικτικές εφαρμογές βιομετρικών συστημάτων

Στο σημείο αυτό θα πρέπει να τονιστεί ότι τα υπάρχοντα βιομετρικά συστήματα είναι κυρίως μικρής κλίμακας με σχετικά περιορισμένες απαιτήσεις και λειτουργούν υπό ποικίλες υποθέσεις. Έτσι, για παράδειγμα, τα έξυπνα (ή ενεργητικά) συστήματα παρακολούθησης τα οποία βασίζονται σε κάμερα και χρησιμοποιούνται σε εσωτερικούς χώρους, όπως μουσεία και χώρους περιορισμένης πρόσβασης, δεν καλούνται να ανταποκριθούν σε συνθήκες μεγάλων μεταβολών

φωτεινότητας, μεταβαλόμενου φόντου ή σε διαφορετικές καιρικές συνθήκες όπως ένα σύστημα εξωτερικού χώρου . Έτσι αυτά τα συστήματα χρησιμοποιούν [58] [59] χρησιμοποιούν σχετικά απλούς και γρήγορους αλγόριθμους αφαίρεσης φόντου και εντοπισμού ανθρώπου σε σχέση με συστήματα εξωτερικού χώρου. Ο όρος «έξυπνα (ή ενεργητικά) συστήματα παρακολούθησης» αναφέρεται σε συστήματα τα οποία μπορούν αυτομάτως να ανιχνεύσουν *ύποπτη* συμπεριφορά (σε σχέση με τη συνήθη, αναμενόμενη) στην περιοχή την οποία καλύπτουν. Τα έξυπνα συστήματα παρακολούθησης, όμως, απαιτούν επεξεργασία της ροής των εικόνων (image stream) σε πραγματικό χρόνο και συνεπώς απαιτείται κατάλληλη ισορρόπηση των απαιτήσεων, με την ισορρόπηση αυτή να είναι κυρίως εξαρτώμενη από την εφαρμογή.

4 *Αρχιτεκτονική Βιομετρικών Συστημάτων*

Στο κεφάλαιο αυτό αναλύονται τα βιομετρικά συστήματα από πλευράς αρχιτεκτονικής. Συγκεκριμένα, παρουσιάζονται οι κύριες μη λειτουργικές τους απαιτήσεις, ενώ εν συνεχεία περιγράφεται αναλυτικά η γενική αρχιτεκτονική των βιομετρικών συστημάτων και τα διάφορα υποσυστήματα αυτής. Τέλος, αναλύονται τα κυριότερα σφάλματα των βιομετρικών συστημάτων καθώς και οι κυριότεροι δείκτες αξιολόγησης αυτών.

4.1 *Μη λειτουργικές απαιτήσεις Βιομετρικών Συστημάτων*

Ένα πρακτικό βιομετρικό σύστημα – πέραν από τα ποιοτικά χαρακτηριστικά του βιομετρικού γνωρίσματος στο οποίο βασίζεται - καλείται να μπορεί να ικανοποιήσει τις προδιαγεγραμμένες απαιτήσεις *ακρίβειας*, *ταχύτητας* και *πόρων*, να είναι *ακίνδυνο* για τους χρήστες, να είναι *αποδεκτό* από τον στοχευόμενο πληθυσμό και να είναι επαρκώς *εύρωστο* σε πληθώρα δόλιων μεθόδων και επιθέσεων προς το σύστημα.

Παράλληλα, αρκετοί παράγοντες πρέπει να ληφθούν υπ’οψιν :

- *Βασικός στόχος* βιομετρικού συστήματος: ταυτοποίηση, επιβεβαίωση ταυτότητας, σύνθεση προφίλ
- Απαιτήσεις χώρου εφαρμογής σε σχέση με την *ασφάλεια* και την *ιδιωτικότητα*
- Το *μέγεθος της βάσης των χρηστών* του συστήματος τα στοιχεία των οποίων θα πρέπει να διαχειρίζονται μέσω αυτού
- Αναγκαιότητα χρήσης του συστήματος σε *πραγματικό χρόνο*
- Αποδεκτό χρηματικό *κόστος*
- Βαθμός *εξοικίωσης* των χρηστών με την τεχνολογία

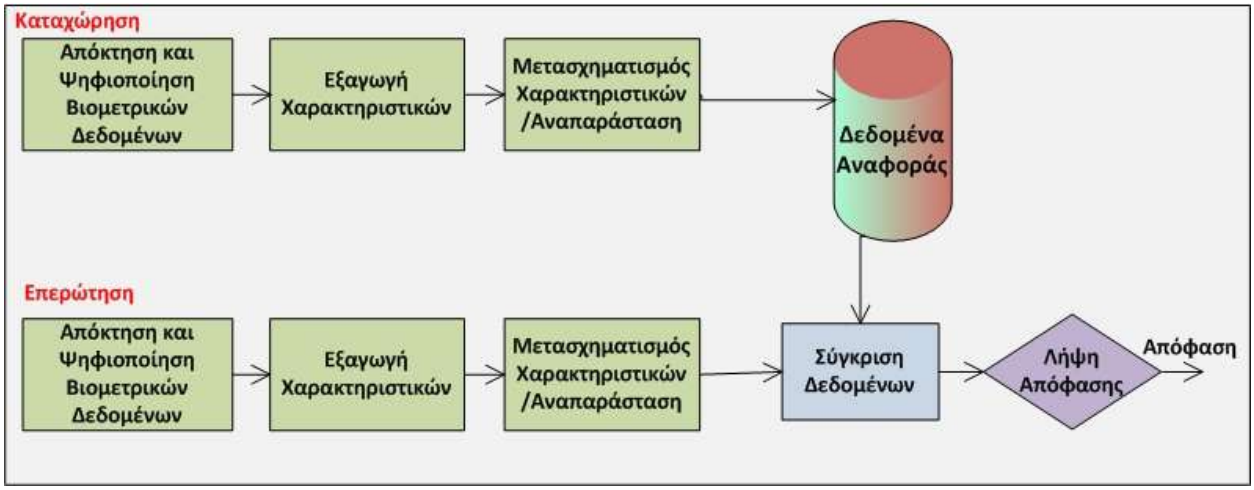
4.2 *Γενική Αρχιτεκτονική Βιομετρικών Συστημάτων*

Για τον προσδιορισμό μιας γενικής αρχιτεκτονικής βιομετρικών συστημάτων και αναλυτικής περιγραφής των επιμέρους δομικών της στοιχείων έχουν λάβει χώρα διάφορες προσπάθειες από οργανισμούς προτυποποίησης, όπως το Biometric Evaluation Methodology (BEM) Working Group και το BioAPI (Application Programming Interface) Consortium. Κάθε μία από αυτές τις προσπάθειες περιλαμβάνει διαφορές σε ορολογίες και μοντέλα, αλλά όλες χαρακτηρίζονται από βασικά κοινά στοιχεία τα οποία επιτρέπουν τη σύνθεση μιας γενικής αρχιτεκτονικής αναφοράς για τα βιομετρικά συστήματα.

4.2.1 Οι δύο φάσεις λειτουργίας των Βιομετρικών Συστημάτων

Όλα τα βιομετρικά συστήματα λειτουργούν σε δύο ξεχωριστές φάσεις; την *καταχώρηση* των βιομετρικών δεδομένων αναφοράς (*enrolment*) και την *επερώτηση* του βιομετρικού συστήματος (Εικόνα 5). Η πρώτη φάση, αποτελεί μέρος της αρχικοποίησης του βιομετρικού συστήματος όταν πια αυτό έχει τεθεί σε λειτουργία. Κατά τη φάση αυτή παρέχονται στο σύστημα δείγματα των βιομετρικών χαρακτηριστικών των ατόμων προκειμένου να παραχθούν τα βιομετρικά

δεδομένα αναφοράς (reference template) μετά την αποθήκευση των οποίων το άτομο θεωρείται γνωστό στο βιομετρικό σύστημα.



Εικόνα 5 Γενικά Στάδια Βιομετρικού Συστήματος

Κατά τη φάση επερώτησης, η οποία αφορά στην κανονική ροή του συστήματος, εισάγονται στο σύστημα νέα βιομετρικά δεδομένα – τα οποία αποτελούν τα δεδομένα επερώτησης. Τα δεδομένα αυτά επεξεργάζονται και συγκρίνονται με τα αποθηκευμένα δεδομένα αναφοράς είτε όλων των καταχωρημένων ατόμων (ταυτοποίηση) είτε ενός συγκεκριμένου χρήστη (επιβεβαίωση ταυτότητας). Αναλόγως με τον κύριο στόχο του συστήματος, το τελευταίο επιστρέφει στην περίπτωση της ταυτοποίησης κάποια στοιχεία ταυτότητας που αφορούν στο συγκεκριμένο άτομο (εφ'όσον έχει βρεθεί στα καταχωρημένα άτομα) ή λίστα με τα πιο πιθανά προς ταυτοποίηση άτομα με βάση την επεξεργασία του συστήματος και στην περίπτωση της επιβεβαίωσης ταυτότητας είτε στοιχεία ταυτότητας για το άτομο του οποίου η ταυτότητα πιστοποιήθηκε είτε ένα απλό ναι/όχι.

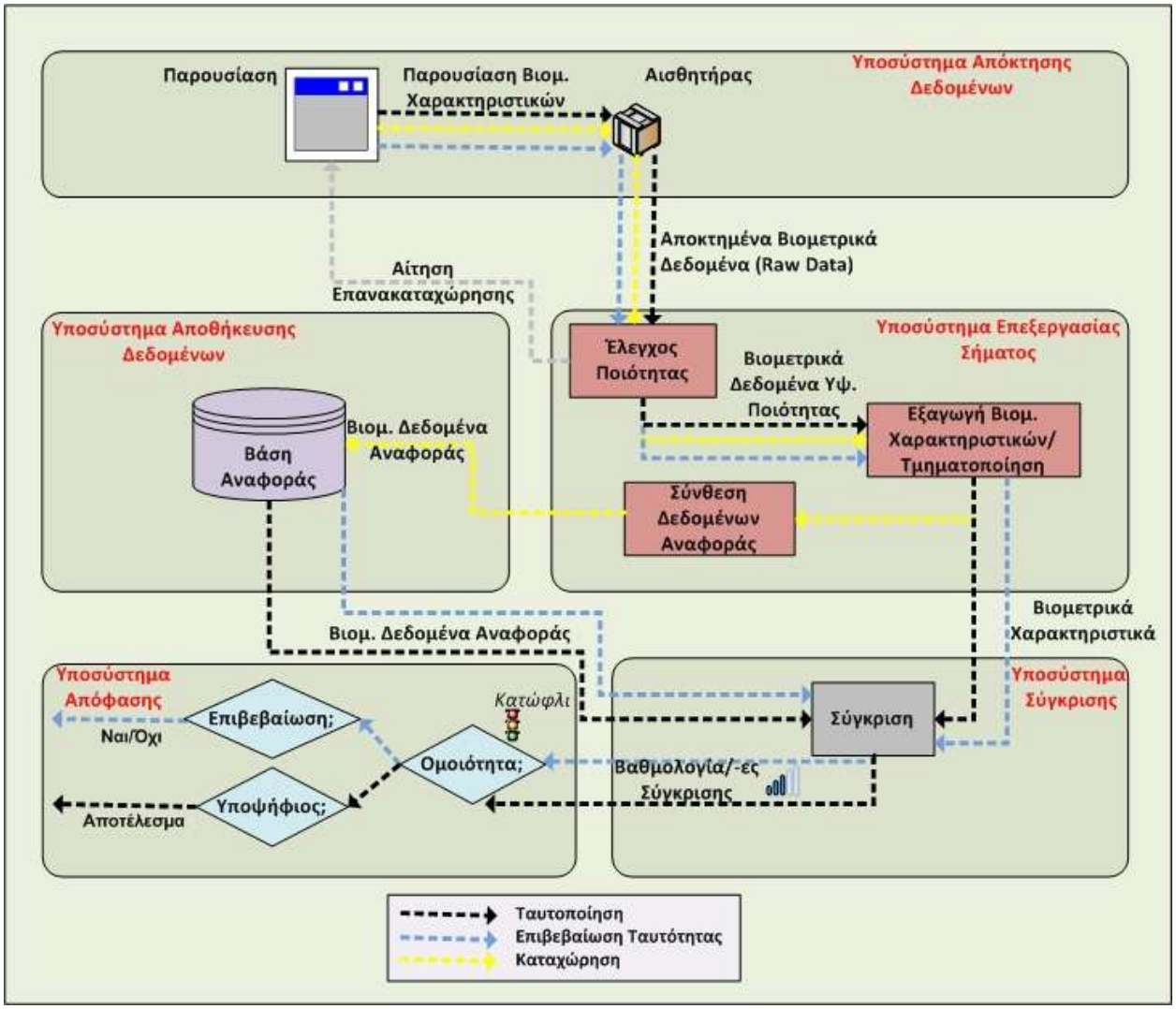
4.2.2 Τα υποσυστήματα ενός γενικού Βιομετρικού Συστήματος

Όπως έχει ήδη αναφερθεί, ένα βιομετρικό σύστημα μπορεί να χρησιμοποιηθεί για ταυτοποίηση, επιβεβαίωση ταυτότητας ή σύνθεση προφίλ ατόμων. Ανεξαρτήτως όμως της

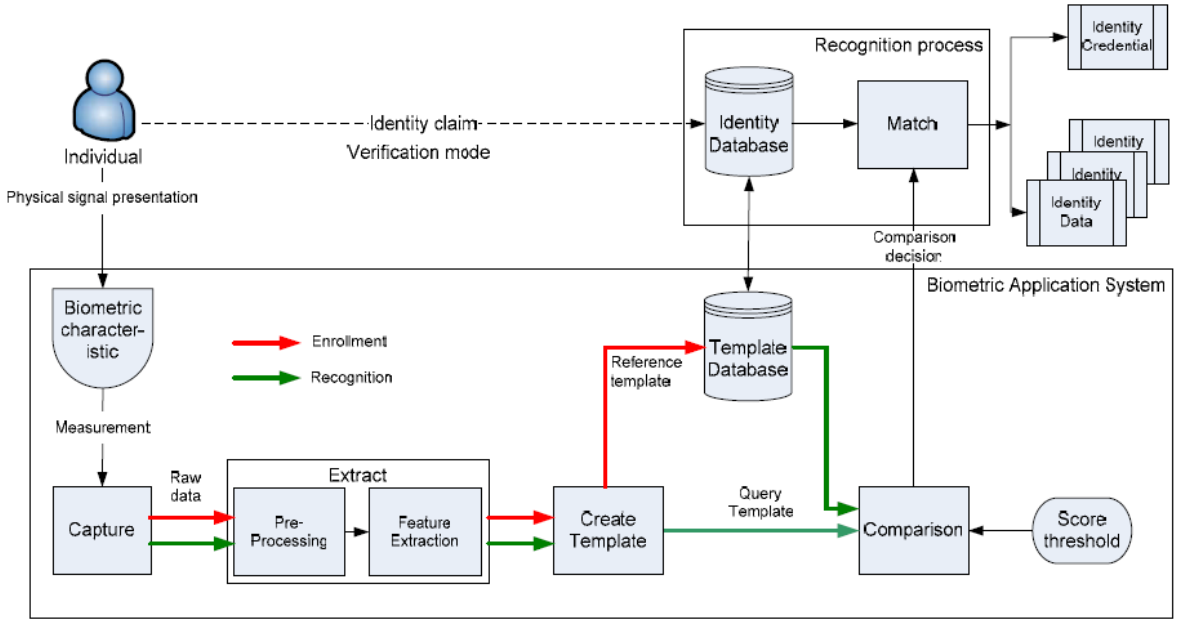
εφαρμογής του συστήματος, όλα τα βιομετρικά συστήματα έχουν κάποιες κοινές λειτουργίες οι οποίες μπορούν να αποτυπωθούν σε μια σειρά από δομικά στοιχεία (Εικόνα 6Εικόνα 6 Γενικό Μοντέλο Βιομετρικού Συστήματος):

- έναν ή περισσότερους *αισθητήρες* μέσω των οποίων αποτυπώνονται τα βιομετρικά χαρακτηριστικά στα οποία εστιάζει το βιομετρικό σύστημα
- μια *αποθηκευτική* μονάδα η οποία περιέχει τα βιομετρικά δεδομένα αναφοράς (reference templates) των χρηστών και με τη σειρά της είναι συνδεδεμένη με μια βάση με τα δεδομένα ταυτότητας των υποκειμένων
- μια διαδικασία *επεξεργασίας* των βιομετρικών δεδομένων τα οποία καταχωρούνται στο σύστημα προκειμένου να γίνει ταυτοποίηση, επιβεβαίωση ταυτότητας ή σύνθεση του προφίλ κατα την οποία αναλόγως με την ακολουθούμενη μέθοδο εξάγονται συγκεκριμένα γνωρίσματα (αποστάσεις, χρώματα, αναλογίες, κορυφές, καμπυλότητα, κλπ) για περαιτέρω ανάλυση
- μια διαδικασία σύγκρισης κατά την οποία αποτιμάται η ομοιότητα μεταξύ των δεδομένων αναφοράς και των καταχωρημένων προς εξέταση δεδομένων προκειμένου να προκύψει ο βαθμός ομοιότητας
- μια διαδικασία απόφασης η οποία λαμβάνοντας τα αποτελέσματα από τη διαδικασία σύγκρισης και με βάση κάποιο κατώφλι ή κάποια πιο εξεζητημένη πολιτική αποφασίζει αν τα καταχωρημένα δεδομένα ταιριάζουν με κάποιο από τα δεδομένα αναφοράς

Αυτές οι βασικές λειτουργίες αποτυπώνονται σε επιμέρους συστήματα των βιομετρικών συστημάτων και αναλύονται στις ακόλουθες παραγράφους.



Εικόνα 6 Γενικό Μοντέλο Βιομετρικού Συστήματος



Εικόνα 7 Εναλλακτικό Γενικό Μοντέλο Βιομετρικού Συστήματος [101]

4.2.2.1 Απόκτηση (Acquisition)

Η λειτουργία αυτή αποτελεί το πρώτο στάδιο ενός βιομετρικού συστήματος κατά το οποίο λαμβάνεται με αυτόματο τρόπο το βιομετρικό δείγμα (φυσιολογικό(-ά) ή συμπεριφορικό(-ά) βιομετρικό(-ά) χαρακτηριστικό(-ά)) από το υπο εξέταση άτομο. Κατά τη φάση αυτή είναι πιθανό να περιλαμβάνονται διαδικασίες οι οποίες βελτιώνουν την ποιότητα του ληφθέντος δείγματος, όπως είναι διεπιφάνεια χρήστη για αναμόρφωση δεδομένων εισόδου (user interface feedback) ή χρήση πολλαπλών λήψεων για καλύτερη λήψη του δείγματος.

Το κύριο δομικό στοιχείο σε αυτό το βήμα είναι ο αισθητήρας (sensor) μέσω του οποίου γίνεται η λήψη και αποτύπωση του βιομετρικού δείγματος τόσο κατά τη φάση της καταχώρησης όσο και κατά τη φάση της επερώτησης του βιομετρικού συστήματος. Παραδείγματα τέτοιων αισθητήρων (biometric scanners) είναι η κάμερα, η συσκευή ανάγνωσης δακτυλικών αποτυπωμάτων (fingerprint reader), το μικρόφωνο και η συσκευή αποτύπωσης της ίριδας (iris scanner). Κάθε συσκευή έχει συγκεκριμένα κριτήρια και διαδικασίες οι οποίες ορίζονται για τη

διαδικασία λήψης του βιομετρικού χαρακτηριστικού. Έτσι, για παράδειγμα, οι συσκευές αναγνώρισης προσώπου μπορεί να απαιτούν το άτομο να βρίσκεται σε όρθια στάση και να κοιτάζει προς την κάμερα, αναλόγως φυσικά και με το είδος της εφαρμογής.

Το στάδιο αυτό είναι κοινό και στις δύο φάσεις των βιομετρικών συστημάτων. Σε αυτό το σημείο θα πρέπει να τονιστεί ότι το περιβάλλον λειτουργίας της συσκευής θα πρέπει να είναι καθορισμένο έτσι ώστε οι περιβαλλοντικές επιρροές να είναι εντός επιτρεπόμενου πλαισίου.

4.2.2.2 Επεξεργασία Σήματος (Signal Processing)

Κατά το στάδιο επεξεργασίας του σήματος το οποίο λαμβάνεται από τη βιομετρική συσκευή, λαμβάνει χώρα εξαγωγή και αρχική επεξεργασία των βιομετρικών χαρακτηριστικών τα οποία ελήφθησαν κατά το προηγούμενο βήμα. Το στάδιο αυτό είναι κρίσιμο από άποψη απόδοσης του συστήματος, καθώς από αυτό εξαρτάται ο βαθμός μοναδικότητας του προτύπου το οποίο θα παραχθεί κατά τη φάση της καταχώρησης με το οποίο θα γίνεται κάθε σύγκριση κατά τη φάση επερώτησης του συστήματος.

Το στάδιο αυτό περιλαμβάνει έλεγχο ποιότητας των ληφθέντων δεδομένων κατά το στάδιο της απόκτηση των βιομετρικών χαρακτηριστικών. Αν η ποιότητα αυτών δεν είναι αποδεκτή, τότε το προηγούμενο στάδιο μπορεί να επαναληφθεί. Συγκεκριμένα και με βάση τα παραπάνω, κατά τη φάση της καταχώρησης των βιομετρικών δεδομένων το απαιτούμενο επίπεδο ποιότητας των ληφθέντων βιομετρικών χαρακτηριστικών είναι αρκετά υψηλό, δεδομένου ότι αυτά θα αποτελέσουν τη βάση για όλες τις βιομετρικές συγκρίσεις κατά τη λειτουργία του συστήματος. Προκειμένου να επιτευχθεί η απαιτούμενη ποιότητα των δεδομένων, πολλαπλές επαναλαμβανόμενες προσπάθειες λήψεις αυτών μπορεί να λάβουν χώρα κατά την καταχώρηση ώστε το καλύτερο δείγμα να χρησιμοποιηθεί ως πρότυπο αναφοράς.

Το στάδιο αυτό είναι κοινό και στις δύο φάσεις των βιομετρικών συστημάτων και αφορά στην επεξεργασία τόσο των βιομετρικών δεδομένων που θα αποτελέσουν το βιομετρικό πρότυπο αναφοράς έναντι του οποίου θα γίνονται οι συγκρίσεις κατά τη φάσης της επερώτησης του συστήματος όσο και των βιομετρικών δεδομένων τα οποία λαμβάνονται κατά τη φάση επερώτησης και στοχεύουν στην ταυτοποίηση ή επιβεβαίωση της ταυτότητας του κατόχου τους. Τα δεδομένα αυτά είναι γνωστά και ως Biometric Live Record (BLR).

4.2.2.3 Δημιουργία προτύπου αναφοράς

Το στάδιο αυτό περιλαμβάνει τη σύνθεση του βιομετρικού προτύπου αναφοράς (template) το οποίο βασίζεται στην έξοδο του προηγούμενου στάδιου, ενώ μπορεί να συμπεριλαμβάνει και την προσθήκη πιστοποιητικών χρήστη (user credentials), την κωδικοποίηση των βιομετρικών δεδομένων και άλλα δεδομένα τα οποία συνδέονται με την ταυτότητα του χρήστη και εξαρτώνται από την εφαρμογή του συστήματος. Το βιομετρικό πρότυπο αναφοράς καλείται και Biometric Identification Record (BIR).

4.2.2.4 Σύγκριση

Κατά το στάδιο αυτό λαμβάνει χώρα η σύγκριση της βιομετρικής πληροφορίας όπως αυτή έχει εξαχθεί από το ληφθέν δείγμα και της βιομετρικής πληροφορίας στο πρότυπο αναφοράς. Αναλόγως με τον στόχο του συστήματος – ταυτοποίηση ή επιβεβαίωση ταυτότητας - η σύγκριση γίνεται με μια λίστα υποψηφίων προτύπων αναφοράς ή με ένα μόνο πρότυπο αναφοράς αντιστοίχως. Η σύγκριση ουσιαστικά λαμβάνει χώρα μεταξύ των βιομετρικών δεδομένων αναφοράς τα οποία και ανακτώνται από τη βάση δεδομένων και του Biometric Live Record του υπό εξέταση υποκειμένου. Είναι σύνηθες η ομοιότητα μεταξύ των δεδομένων αυτών να

εκφράζεται μέσω κάποιας τιμής (score) και όσο μεγαλύτερη είναι αυτή η τιμή, τόσο μεγαλύτερη να είναι η ομοιότητα μεταξύ τους.

Έτσι, το αποτέλεσμα του στάδιου αυτού είναι μια τιμή που δείχνει πόσο μοιάζουν τα συγκρινόμενα δεδομένα στην περίπτωση της επιβεβαίωσης ταυτότητας και μια λίστα με υποψήφια άτομα τα βιομετρικά πρότυπα αναφοράς των οποίων ταιριάζουν «καλύτερα» - το match score είναι υψηλότερο – σε σχέση με των υπολοίπων καθώς και το υπολογισμένο match score στην περίπτωση της ταυτοποίησης.

4.2.2.5 Απόφαση

Το στάδιο της απόφασης λαμβάνει ως είσοδο το αποτέλεσμα της σύγκρισης και με βάση ένα προκαθορισμένο κατώφλι (threshold) προχωρά σε καθορισμό της απόφασης του συστήματος. Το κατώφλι αυτό είτε είναι καθορίσιμο από τον διαχειριστή του βιομετρικού συστήματος είτε είναι σταθερό. Είναι σημαντικό να τονιστεί ότι δεδομένης της σπουδαιότητας του κατωφλιού αυτού για τον καθορισμό της τελικής απόφασης σημαντικά επίπεδα ασφαλείας απαιτούνται προκειμένου η αλλαγή του – αν είναι δυνατή – να γίνεται μόνο από εξουσιοδοτημένα άτομα. Όπως έχει ηδη αναφερθεί, η τιμή αυτού του κατωφλιού εξαρτάται από τον στόχο του συστήματος και το πεδίο εφαρμογής του.

Έτσι, αν το σύστημα ταυτοποιήσει το υποκείμενο ή επιβεβαιώση την ταυτότητά του, τότε τα πιστοποιητικά του ατόμου, πληροφορίες σχετικές με αυτό και άλλα παρέχονται από το σύστημα. Συγκεκριμένα στην περίπτωση της επιβεβαίωσης ταυτότητας, η τελική απόφαση μπορεί να απαιτεί την προσκόμιση ή υποβολή και επιπρόσθετων στοιχείων, όπως ενός κωδικού ή μιας έξυπνης κάρτας (smart card). Επιπροσθέτως, η διαδικασία της σύγκρισης μπορεί να περιλαμβάνει περισσότερα από ένα βιομετρικά χαρακτηριστικά. Σε τέτοια πολυτροπικά

συστήματα, η τελική απόφαση συνήθως εξαρτάται από μια μαθηματική διαδικασία η οποία συνδυάζει τα αποτελέσματα των διαφόρων διαδικασιών σύγκρισης.

4.2.2.5.1 Το κατώφλι

Θεωρητικά η τιμή του αποτελέσματος σύγκρισης των βιομετρικών δεδομένων του υπό εξέταση ατόμου (client score) θα πρέπει να είναι πάντα μεγαλύτερη από τις τιμές των υπολοίπων ατόμων τα οποία είναι καταχωρημένα στο σύστημα. Αν αυτό ίσχυε πάντα, τότε η επιλογή ενός κατωφλιού το οποίο να διαχωρίζει αυτές τις δύο ομάδες (clients versus impostors) θα αρκούσε για τη λειτουργία του συστήματος. Δεδομένου όμως ότι αυτό δεν ισχύει κατά την πρακτική εφαρμογή των βιομετρικών συστημάτων για διάφορους λόγους, τα βιομετρικά συστήματα συνοδεύονται από μια τιμή σφάλματος. Οι λόγοι αυτοί, μη λαμβάνοντας υπ'όψιν προσπάθειες κλοπής ή αντιγραφής της ταυτότητας ενός ατόμου στο βιομετρικό σύστημα, συμπεριλαμβάνουν ανεπαρκή ποιότητα του ληφθέντος δείγματος βιομετρικών χαρακτηριστικών, εγγενείς αδυναμίες των αλγορίθμων εξαγωγής βιομετρικών χαρακτηριστικών και των αλγορίθμων κατηγοριοποίησης και σύγκρισης (classification algorithms), υπάρχουσες συνθήκες κατά τη λήψη των δεδομένων (περιβάλλον, θόρυβος) και άλλους.

4.3 *Είδη Σφαλμάτων Βιομετρικών Συστημάτων*

Τα συστηματικά και στατιστικά σφάλματα των μετρήσεων και των αλγόριθμων των διαδικασιών εξαγωγής των βιομετρικών χαρακτηριστικών και σύγκρισης καθορίζουν τα όρια της εφαρμογής ενός βιομετρικού συστήματος και την ικανότητα διαχωρισμού μεταξύ διαφορετικών ατόμων. Μια σειρά από μέτρα τα οποία έχουν οριστεί προκειμένου να αποτιμήσουν την απόδοση των βιομετρικών συστημάτων. Τα μέτρα αυτά αφορούν σε είδη σφαλμάτων τα οποία μπορεί να λάβουν χώρα όσον αφορά στην εκτίμηση του βιομετρικού

συστήματος. Έτσι, έχουν καθοριστεί δύο είδη σφαλμάτων: τα *σφάλματα σύγκρισης* (matching errors) και τα *σφάλματα απόκτησης* (acquisition errors).

Το σύνηθες αποτέλεσμα ενός βιομετρικού συστήματος κατά το στάδιο της σύγκρισης είναι το λεγόμενο matching score S, το οποίο αποτελεί ένα μέτρο ομοιότητας μεταξύ των βιομετρικών δεδομένων που υποβλήθηκαν κατά τη φάση της επερώτησης και των βιομετρικών δεδομένων τα οποία αποθηκεύτικα ως πρότυπο αναφοράς κατά τη φάση της καταχώρησης. Το μέγεθος αυτό συγκρίνεται με ένα κατώφλι T προκειμένου να ληφθεί η απόφαση για το αν τα δύο δείγματα προέρχονται από το ίδιο βιομετρικό χαρακτηριστικό ή όχι. Έτσι, όπως έχουμε δει, αν $S \geq T$ (υποθέτωντας ότι αν πρόκειται για διαδικασία ταυτοποίησης τα matching scores είχαν ταξινομηθεί κατά φθίνουσα σειρά με το μεγαλύτερο να αντιπροσωπεύει μεγαλύτερο βαθμό ομοιότητας με το πρότυπο αναφοράς), τότε τα δύο δείγματα θεωρούνται όμοια, ενώ στην αντίθετη περίπτωση θεωρούνται ανόμοια. Η επιλογή του κατωφλιού είναι πολή σημαντική για το ποσοστό των εσφαλμένων ανόμοιων (σφάλμα τύπου 1) ή για το ποσοστό εσφαλμένων όμοιων (σφάλμα τύπου 2) , τα οποία και θα αναλυθούν στις ακόλουθες παραγράφους.

Έστω ότι το αποθηκευμένο πρότυπο αναφοράς είναι R και τα ληφθέντα βιομετρικά δεδομένα επερώτησης είναι Q, τότε έχουμε τις δύο ακόλουθες υποθέσεις:

$$H_0 : Q \approx R \text{ και } H_1 : Q \neq R$$

και αντιστοίχως τις αποφάσεις D_0, που συμβολίζει την απόφαση ότι τα υποβεβλημένα βιομετρικά δεδομένα προέρχονται από το ίδιο βιομετρικό χαρακτηριστικό (κι επομένως το ίδιο άτομο) με το πρότυπο αναφοράς και D_1, που συμβολίζει την απόφαση ότι τα ληφθέντα βιομετρικά δεδομένα δεν προέρχονται από το ίδιο βιομετρικό χαρακτηριστικό με το πρότυπο αναφοράς [101].

Με βάση τα παραπάνω μπορούν να ληφθούν τέσσερεις αποφάσεις:

a) *Αληθώς Αρνητικό* (True Negative - TN): όταν λήφθηκε η απόφαση D_1, και το H_1 είναι αληθές. Στην περίπτωση αυτή, το σήμα δεν ανήκει στο άτομο X και ισχυριζόμαστε ότι δεν είναι το άτομο X.

b) *Εσφαλμένως Θετικό* (False Positive - FP): όταν λήφθηκε η απόφαση D_0, και το H_1 είναι αληθές. Στην περίπτωση αυτή, το σήμα δεν ανήκει στο άτομο X και ισχυριζόμαστε ότι το άτομο αυτό είναι είναι το άτομο X.

c) *Εσφαλμένως Αρνητικό* (False Negative - FN): όταν λήφθηκε η απόφαση D_1, και το H_0 είναι αληθές. Στην περίπτωση, το σήμα ανήκει στο άτομο X και ισχυριζόμαστε ότι δεν είναι το άτομο X.

d) *Αληθώς Θετικό* (True Positive - TP): όταν λήφθηκε η απόφαση D_0, και το H_0 είναι αληθές. Στην περίπτωση αυτή, το σήμα ανήκει στο άτομο X και ισχυριζόμαστε ότι είναι το άτομο X.

Όσον αφορά στα σφάλματα σύγκρισης αυτά προσδιορίζονται από τον *Δείκτη Λανθασμένης Αποδοχής* (*False Acceptance Rate-FAR*) και τον *Δείκτη Λανθασμένης Απόρριψης* (*False Rejection Rate-FRR*). Ο πρώτος δείκτης αναφέρεται στην εσφαλμένη ταυτοποίηση ή επιβεβαίωση ταυτότητας ενός ατόμου καθώς το σύστημα αναγνωρίζει διαφορετικά βιομετρικά χαρακτηριστικά ως πανομοιότυπα και θεωρείται ότι αποτελεί το πιο σημαντικό σφάλμα ασφαλείας ενός βιομετρικού συστήματος. Με άλλα λόγια, είναι ένα μέτρο της πιθανότητας το σύστημα να επιτρέψει λανθασμένα πρόσβαση σε ένα μη εξουσιοδοτημένο άτομο. Έτσι, αν ο δείκτης αυτός είναι 10%, τότε κατά μέσο όρο ένα στα δέκα μη εξουσιοδοτημένα άτομα τα οποία προσπαθούν να αποκτήσουν πρόσβαση σε ένα σύστημα θα αναγνωριστούν από το σύστημα ως εξουσιοδοτημένα. Με βάση τα παραπάνω ο δείκτης αυτός μπορεί να οριστεί ως το πηλίκο των

Εσφαλμένων Θετικών προς το σύνολο των περιπτώσεων που το σήμα δεν ανήκει στο άτομο ως εξής:

$$FAR = \frac{FP}{TN + FP}$$

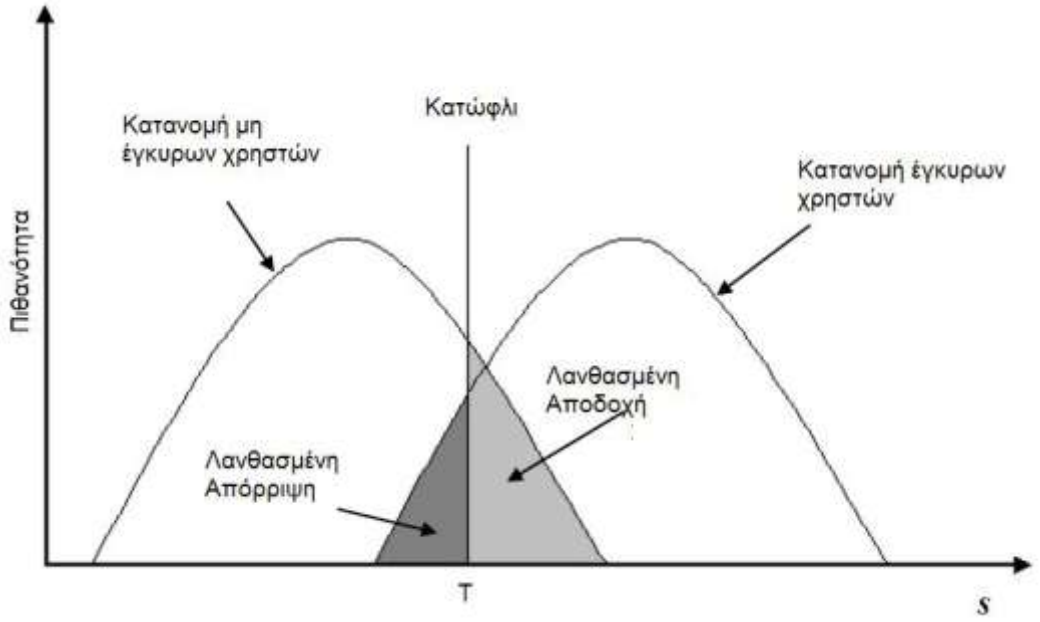
Ο *Δείκτης Λανθασμένης Απόρριψης* αφορά στην αποτυχία του συστήματος να ταυτοποιήσει ή να επιβεβαιώση την ταυτότητα ενός ατόμου. Στην περίπτωση αυτή, πανομοιότυπα χαρακτηριστικά αναγνωρίζονται από το σύστημα ως διαφορετικά. Αν και ο δείκτης αυτός δεν είναι τόσο κρίσιμος όσο ο προηγούμενος, εντούτοις υψηλές τιμές του μπορεί να οδηγήσουν σε σημαντική μείωση της αξιοπιστίας του συστήματος και συνακολούθως σε μείωση της αποδοχής του και σε ενοχλημένους χρήστες. Ομοίως ο δείκτης αυτός μπορεί να οριστεί ως το πηλίκο των Εσφαλμένων Αρνητικών προς το σύνολο των περιπτώσεων που το σήμα ανήκει στο άτομο ως εξής:

$$FRR = \frac{FN}{TP + FN}$$

Οι παραπάνω δείκτες με βάση τον ορισμό τους αφορούν σε συστήματα επιβεβαίωσης ταυτότητας. Δύο γενικότεροι δείκτες είναι ο Δείκτης Λανθασμένης Ομοιότητας (False Match Rate (FMR)) και ο Δείκτης Λανθασμένης Ανομοιότητας (False Non-Match Rate (FNMR), οι οποίοι ορίζονται ομοίως με τον Δείκτη Λανθασμένης Αποδοχής και τον Δείκτη Λανθασμένης Απόρριψης αντιστοίχως, αλλά αφορούν τόσο σε συστήματα ταυτοποίησης όσο και επιβεβαίωσης ταυτότητας.

Στην ακόλουθη εικόνα παρουσιάζονται οι τυπικές κατανομές των score παραμέτρων σύγκρισης μεταξύ προτύπων αναφοράς και ληφθέντων βιομετρικών χαρακτηριστικών τόσο από το άτομο στο οποίο ανήκουν και τα δεδομένα πρότυπου αναφοράς (πράσινη γραμμή) όσο και τρίτων (κόκκινη γραμμή). Το πράσινο πεδίο δεξιά της τιμής του κατωφλιού αντιπροσωπεύει το

συνολικό FNMR και το κόκκινο πεδίο δεξιά της τιμής του κατωφλιού αντιπροσωπεύει το συνολικό FMR. Τα δύο σημεία ZFR και ZFA αφορούν τις κρίσιμες τιμές του score κάτω από τις οποίες το FNMR γίνεται μηδενικό και πάνω από τις οποίες το FMR γίνεται μηδενικό αντιστοίχως.



Εικόνα 8 Κατανομή match scores τρίτων και χρηστών συστήματος και δείκτες λανθασμένης αποδοχής και λανθασμένης απόρριψης [101]

Ακολουθώντας μια πιο επίσημη αναπαράσταση των δεικτών FMR και FNMR με βάση τις υποθέσεις H_0 και H_1 οι οποίες ορίστηκαν παραπάνω προκύπτει ότι για συστήματα επιβεβαίωσης ταυτότητας:

$$\mathrm{FMR(T)} = \int_T^1 p(S|\mathrm{H}_1 \text{is true})\mathrm{dS}$$

που δείχνει τη συνολική πιθανότητα το υπολογισμένο score να έχει τιμή μεγαλύτερη από το κατώφλι αν και τα δεδομένα του πρότυπου αναφοράς και τα ληφθέντα βιομετρικά δεδομένα δεν προέρχονται από το ίδιο υποκείμενο και

$$\mathrm{FNMR(T)} = \int_0^T p(S|\mathrm{H}_0 \text{is true})\mathrm{dS}$$

που δείχνει τη συνολική πιθανότητα το υπολογισμένο score να έχει τιμή μικρότερη από το κατώφλι αν και τα δεδομένα του πρότυπου αναφοράς και τα ληφθέντα βιομετρικά δεδομένα προέρχονται από το ίδιο υποκείμενο.

Στην περίπτωση συστημάτων πιστοποίησης ταυτότητας η διαδικασία της σύγκρισης χρησιμοποιεί όλα τα πρότυπα αναφοράς (έστω R_i) και περιλαμβάνει τον υπολογισμό όλων των matching scores (έστω $S_i(Q, R_i)$) των ληφθέντων βιομετρικών χαρακτηριστικών με τα πρότυπα αναφοράς. Στην περίπτωση αυτή το αποτέλεσμα, όπως έχουμε αναφέρει, γενικώς είναι μια ταξινομημένη λίστα από πρότυπα αναφοράς τα οποία έχουν matching score πάνω από την τιμή του κατωφλιού, δηλαδή όλα τα $S_j > T$, $j < N$, με N το πλήθος των καταχωρημένων βιομετρικών προτύπων αναφοράς στο σύστημα. Όσο μεγαλύτερη είναι η αρχική βάση προτύπων αναφοράς τόσο μεγαλύτερη είναι γενικώς η πιθανότητα το μέγεθος της λίστας να είναι μεγαλύτερο καθώς και η πιθανότητα να υπάρχει ένα λάθος. Έτσι, η λήψη της απόφασης γίνεται πιο περίπλοκη καθώς δεν είναι δεδομένο ποιο από τα αποτελέσματα σύγκρισης που είναι συμπεριλαμβανόμενο μέσα στη λίστα ανήκει στο ίδιο άτομο.

Διακρίνουμε τις ακόλουθες περιπτώσεις:

- Για m καταχωρημένα πρότυπα αναφοράς προέκυψε : $S_i(Q, R_i) < T$, και για τα υπόλοιπα N-m καταχωρημένα πρότυπα $S_j(Q, R_j) > T$ μεταξύ των οποίων είναι και το $S_c(Q, R_c) > T$, δηλαδή η διαδικασία ταυτοποίησης έδωσε ως αποτέλεσμα μια λίστα με υποψήφιους στην οποία περιλαμβάνονται τα βιομετρικά χαρακτηριστικά του κατόχου,του οποίου η πιθανότητα είναι:

$$P_1 = \binom{N-1}{m}(1 - FMR(T))^m \, FMR(T)^{N-1-m} \, (1 - FNMR(T))$$

Στην ειδική περίπτωση κατά την οποία m=N-1, η διαδικασία της ταυτοποίησης λειτούργησε σωστά και έδωσε ως αποτέλεσμα τον πραγματικό κάτοχο των βιομετρικών δεδομένων και μόνο, με την πιθανότητα αυτής να είναι:

$$P_2 = (1 - FMR(T))^{N-1}(1 - FNMR(T))$$

- Για m καταχωρημένα πρότυπα αναφοράς προέκυψε : $S_i(Q, R_i) < T$ μεταξύ των οποίων είναι και το $S_c(Q, R_c) < T$, και για N-m καταχωρημένα πρότυπα $S_j(Q, R_j) > T$, δηλαδή η διαδικασία ταυτοποίησης έδωσε ως αποτέλεσμα μια λίστα με υποψήφιους η οποία δεν περιλαμβάνει τα βιομετρικά χαρακτηριστικά του ατόμου στο οποίο ανήκουν,του οποίου η πιθανότητα είναι:

$$P_3 = \binom{N-1}{m-1}(1 - FMR(T))^{m-1}\, FMR(T)^{N-m}\, FNMR(T))$$

Στην ειδική περίπτωση κατά την οποία m=N, η διαδικασία της ταυτοποίησης δεν επέστρεψε κανένα θετικό αποτέλεσμα, το οποίο μπορεί να οφείλεται είτε στο γεγονός ότι το άτομο του οποίου τα βιομετρικά δεδομένα υποβλήθηκαν για επερώτηση στο βιομετρικό σύστημα δεν είναι καταχωρημένα, είτε στο ότι κατά τη σύγκριση των ληφθέντων βιομετρικών χαρακτηριστικών με το σωστό πρότυπο αναφοράς το προκύπτον matching score είχε τιμή μικρότερη από το κατώφλι. Αν θεωρήσουμε ότι το άτομο ήταν καταχωρημένο στο σύστημα, τότε η πιθανότητα να συμβεί αυτό είναι :

$$P_4 = (1 - FMR(T))^{N-1}FNMR(T)$$

Με βάση την παραπάνω ανάλυση και θεωρόντας τα αποτελέσματα που αφορούν στην πρώτη περίπτωση ως true positive ή correct match και τα αποτελέσματα που αφορούν στη δεύτερη περίπτωση ως false negative ή false rejection τότε συνδυάζοντας τις παραπάνω σχέσεις λαμβάνουμε ότι :

$$FMR_{id}(T) = FNMR(T) \sum_{m=1}^{N-1} \binom{N-1}{m-1} (1 - FMR(T))^{m-1} FMR(T)^{N-m} \Rightarrow$$

$$FMR_{id}(T) = FNMR(T) \left(1 - (1 - FMR(T))^{N-1}\right)$$

και

$$FNMR_{id}(T) = FNMR(T) \sum_{m=1}^{N} \binom{N-1}{m-1} (1 - FMR(T))^{m-1} FMR(T)^{N-m} \Rightarrow$$

$$FNMR_{id}(T) = FNMR(T)$$

Είναι εμφανές ότι και οι δύο αυτές παράμετροι θα πρέπει να έχουν όσο το δυνατόν μικρότερες τιμές αν είναι επιθυμητό το σύστημα να προσφέρει κατάλληλα επίπεδα ασφάλειας και να αποφεύγονται περιπτώσεις ενοχλημένων χρηστών εξαιτίας επαναλαμβανόμενης ή ακανόνιστης απόρριψης. Στο σημείο αυτό θα πρέπει να τονιστεί ότι αν και η διεξαγωγή συγκριτικών εκτιμήσεων για την παροχή σχετικών ενδείξεων είναι δυνατή, τα στοιχεία για τις συνθήκες υπό τις οποίες κάθε σύνολο από τιμές των δεικτών αυτών προέκυψε πολύ συχνά δεν είναι γνωστά και για τον λόγο αυτόν η πραγματική σύγκριση αυτών αποτελεί ιδιαιτέρως δύσκολη διαδικασία. Έτσι, αν και διάφοροι βιομετρικοί αισθητήρες υπόσχονται εξαιρετικά χαμηλές τιμές του FAR – της τάξης του 0.0001% - η εφαρμογή αυτών στην πράξη καταδεικνύει ότι οι τιμές αυτές απέχουν αρκετά από τις παρατηρούμενες. Αυτές οι αποκλίσεις οφείλονται σε μια πληθώρα παραγόντων. Μεταξύ των άλλων, είναι εξαιρετικά δύσκολη – συχνά ακατόρθωτη - και χρονοβόρα διαδικασία να πραγματοποιηθούν δοκιμές σε ένα ευρύ πληθυσμιακό φάσμα το οποίο να ενσωματώνει κάθε είδους ποικιλομορφία η οποία μπορεί να απαντηθεί κατά τη χρήση του συστήματος. Επίσης, τα αποτελέσματα εξαρτώνται από περιβαλλοντολογικές συνθήκες, όπως είναι η θέση του χρήστη, ο τύπος του χρήστη καθώς και οι ρυθμίσεις του επιπέδου ασφάλειας. Σε τεχνικό επίπεδο, τα βιομετρικά συστήματα καλούνται να «αντιμετωπίσουν» θορυβώδη δεδομένα από τους αισθητήρες, περιορισμένους βαθμούς ελευθερίας καθώς και την πιθανότητα ύπαρξης πανομοιότυπων βιομετρικών χαρακτηριστικών μεταξύ ατόμων του πληθυσμού. Παρατηρούμε

επίσης ότι ακόμη και μια πολύ μικρή τιμή του FMR οδηγεί σε πολύ μεγάλες τιμές του FMR_{id} για μεγάλες τιμές του N, με αποτέλεσμα η απόδοση των βιομετρικών τεχνολογιών να πέφτει κατακόρυφα σε μεγάλης κλίμακας εφαρμογές, οι οποίες περιλαμβάνουν μεγάλο πλήθος βιομετρικών δεδομένων προς σύγκριση.

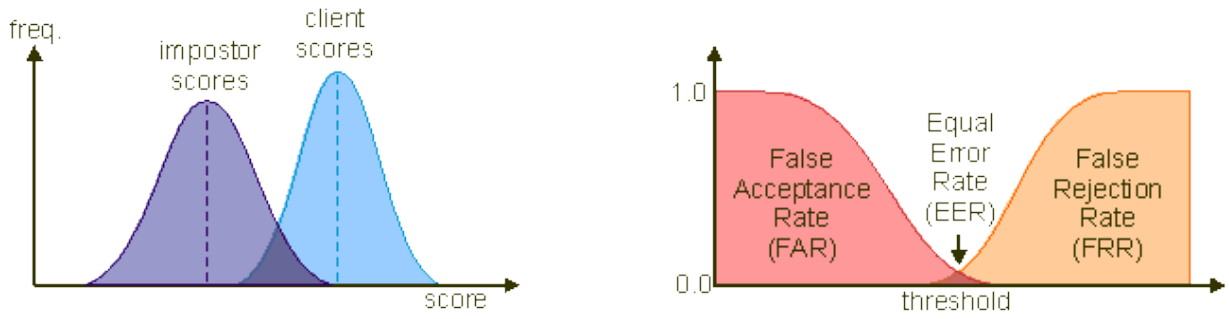
Θα πρέπει να τονιστεί ότι οι δύο δείκτες αυτοί κρύβουν έναν εγγενή περιορισμό καθώς είναι αμοιβαίως αποκλειόμενοι – αν η τιμή του ενός ανέβει, τότε η τιμή του άλλου θα μειωθεί και αντιστρόφως [15] . Για τον λόγο αυτόν, πολλά από τα διαθέσιμα βιομετρικά συστήματα δείχνουν ανοχή σε ιδιαιτέρως υψηλές τιμές του δείκτη FRR προκειμένου να κρατήσουν σε χαμηλές τιμές τον δείκτη FAR. Για πολλές βιομετρικές εφαρμογές αυτό δεν αποτελεί ιδιαίτερο πρόβλημα καθώς είναι συχνά πιθανό να βρεθεί η κατάλληλη ισορροπία μεταξύ της ευαισθησίας σε περιπτώσεις λανθασμένης αποδοχής και λανθασμένης απόρριψης. Όμως, δεν είναι λίγοι οι τομείς που έχουν ιδιαίτερη ευαισθησία και στους δύο δείκτες, όπως είναι για παράδειγμα ο οικονομικός τομέας στον οποίο η εφαρμογή των βιομετρικών συστημάτων εστιάζει στον έλεγχο πρόσβασης για την πραγματοποίηση συναλλαγών οικονομικού χαρακτήρα.

Αυτός ο εγγενής περιορισμός μεταξύ των δύο δεικτών έχει άμεσες επιπτώσεις στο επιλεγόμενο κατώφλι κατά το στάδιο της σύγκρισης. Έτσι, για παράδειγμα, αν επιλεγεί ένα κατώφλι αρκετά υψηλό έτσι ώστε να μην είναι δυνατό το match score ενός άλλου ατόμου να το ξεπεράσει και να αποφευχθούν οι περιπτώσεις λανθασμένης αποδοχής από το σύστημα. Όμως τότε στις περιπτώσεις που το match score του υπό εξέταση ατόμου είναι χαμηλότερο από το κατώφλι αυτό, το άτομο αυτό δε θα ταυτοποείται ορθώς από το σύστημα. Στην αντίθετη περίπτωση που το κατώφλι επιλεγεί να είναι τόσο χαμηλό ώστε να αποφεύγεται κάθε λανθασμένη απόρριψη, τότε αυξάνονται οι πιθανότητες λανθασμένης αποδοχής. Η επιλογή ενός κατωφλιού ανάμεσα σε αυτά τα δύο σημεία οδηγεί σε περιπτώσεις τόσο λανθασμένης αποδοχής

όσο και λανθασμένης απόρριψης. Προκειμένου αυτό να γίνει καλύτερα αντιληπτό ακολουθεί ένα παράδειγμα. Έστω ότι λαμβάνει χώρα δοκιμή ενός βιομετρικού συστήματος επιβεβαίωσης ταυτότητας κάνοντας χρήση μεγάλου αριθμού δοκιμαστικών δεδομένων, τα οποία αποτελούνται τόσο από δεδομένα χρηστών του συστήματος όσο και από τρίτων.

Στην ακόλουθη εικόνα (Εικόνα 9) παρατίθενται τόσο οι κατανομές των match scores των τρίτων όσο και των χρηστών του συστήματος. Στο παράδειγμα αυτό έχει επιλεγεί κατανομή Gauss. Όπως μπορούμε να παρατηρήσουμε, η μέση τιμή της κατανομής των χρηστών του συστήματος είναι μεγαλύτερη από τη μέση τιμή των τρίτων που προσπαθούν να «ξεγελάσουν το σύστημα». Η επιλογή του κατωφλιού καθίσταται δύσκολη όταν οι δύο κατανομές είναι επικαλυπτόμενες.

Αναλόγως με την επιλογή του κατωφλιού, η λανθασμένη αποδοχή μπορεί να λάβει χώρα για όλους, μερικούς ή κανέναν τρίτο. Όπως προκύπτει και από τον ορισμό του Δείκτη Λανθασμένης Αποδοχής, αν η τιμή του ισούται με τη μονάδα, αυτό μεταφράζεται στο γεγονός ότι όλοι οι τρίτοι που προσπαθούν να «ξεγελάσουν» το βιομετρικό σύστημα γίνονται αποδεκτοί από το τελευταίο. Ομοίως στην περίπτωση του Δείκτη Λανθασμένης Απόρριψης, του οποίου η τιμή κυμαίνεται από 0 ως 1, όσο μεγαλύτερη είναι η τιμή του τόσο περισσότεροι χρήστες του συστήματος λανθασμένα απορρίπτονται από το σύστημα.



Εικόνα 9 a) Κατανομή match scores τρίτων και χρηστών συστήματος b) Γραφική παράσταση δεικτών λανθασμένης αποδοχής και λανθασμένης απόρριψης (η τομή τους δίνει το Equal Error Rate)

Δύο ακόμη μέτρα που χρησιμοποιούνται για την αξιολόγηση των βιομετρικών συστημάτων και αφορούν σε σφάλματα απόκτησης είναι ο Δείκτης Εσφαλμένης Καταχώρησης (Failure to Enrol Rate-FER) και ο Δείκτης Ίσου Σφάλματος (Equal Error Rate-EER). Ο FER αναφέρεται στην ικανότητα του συστήματος να καταχωρήσει ένα βιομετρικό χαρακτηριστικό για έναν χρήστη και υποδηλώνει την πιθανότητα το σύστημα να μην είναι σε θέση να εξαγάγει ευδιάκριτα, συνεπή και αντιγράψιμα χαρακτηριστικά από ένα δείγμα το οποίο παρουσιάζεται κατά τη διαδικασία εκχώρησης (enrollment process). Με άλλα λόγια, αφορά στην πιθανότητα το σύστημα να μην καταφέρει να δημιουργήσει δεδομένα αναφοράς για έναν νέο χρήστη. Οι λόγοι πίσω από τέτοιου είδους σφάλματα μπορεί να είναι συμπεριφορικοί, όπως ο χρήστης να κινείται κατά τη διάρκεια της καταχώρησης των δεδομένων, ή και και για σωματικούς λόγους, όπως είναι δυσδιάκριτα ή άτονα πρότυπα (patterns) λόγω ρούχων ή ηλικίας.

Ο EER αναφέρεται στο σημείο τομής των δύο δεικτών FRR και FAR, δηλαδή στο σημείο για το οποίο για συγκεκριμένη τιμή του κατωφλιού ισχύει FRR=FAR ή FMR = FNMR, αναλόγως με τους δείκτες οι οποίοι χρησιμοποιούνται. Έτσι, με βάση τα παραπάνω:

$$\text{EER} = \int_{T}^{1} p(S|\text{H}_1 \text{is true})\text{dS} = \int_{0}^{T} p(S|\text{H}_0 \text{is true})\text{dS}$$

Ο δείκτης αυτός είναι γνωστός και ως *Crossover Error Rate (CER)*. Όσο πιο μικρή είναι η τιμή αυτού του δείκτη, τόσο πιο υψηλή είναι η ακρίβεια του βιομετρικού συστήματος.

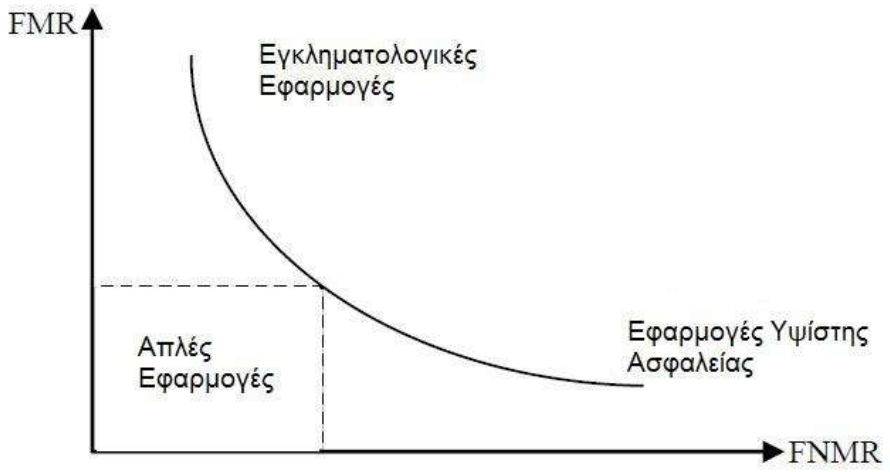
Στο σημείο αυτό θα πρέπει να τονιστεί ότι ένα βιομετρικό σύστημα θα πρέπει να είναι εξαρτώμενο από την εφαρμογή. Έτσι, για παράδειγμα, ένα σύστημα αναγνώρισης δαχτυλικών αποτυπωμάτων το οποίο είναι σχεδιασμένο για ένα περιβάλλον με υψηλές απαιτήσεις ασφάλειας θα έχει διαφορετικές προδιαγραφές ως προς τους δείκτες που παρουσιάστηκαν σε σχέση με ένα αντίστοιχο βιομετρικό σύστημα για έλεγχο πρόσβασης σε έναν υπολογιστή. Στην πρώτη περίπτωση, η χαμηλή τιμή του FAR είναι πολύ πιο σημαντική από μια χαμηλή τιμή του FRR, κι επομένως το κόστος για τον FAR είναι πολύ μεγαλύτερο καθώς εσφαλμένη αναγνώριση ατόμου μπορεί να προσφέρει πρόσβαση σε επικίνδυνο άτομο. Αντιθέτως στη δεύτερη περίπτωση, η ευκολία αποτελεί πιο σημαντικό θέμα. Έτσι, οι χρήστες θα ενοχληθούν σημαντικά αν δεν είναι σε θέση να συνδεθούν στον υπολογιστή τους. Από τα παραπάνω συμπεραίνουμε ότι το κόστος αποτελεί μια πολύ σημαντική παράμετρο η οποία και θα πρέπει να λαμβάνεται υπ'όψιν κατά την ανάλυση της απόδοσης ενός βιομετρικού συστήματος.

Προκειμένου, λοιπόν, να χαρακτηριστεί ένα βιομετρικό σύστημα, η αλληλεξάρτηση των δύο δεικτών μπορεί να αναπαρασταθεί με την καμπύλη Λειτουργικού Χαρακτηριστικού Λήπτη (Receiver Operating Characteristic - ROC). Η καμπύλη αυτή δείχνει τον FNMR (ή τον FRR) συναρτήσει του FMR (ή του FAR αντιστοίχως) και η γραφική παράσταση συχνότατα αναπαρίσταται σε ένα διπλό λογαριθμικό διάγραμμα, ενώ μερικές φορές ο κάθετος άξονας αναπαριστά την τιμή του (1-FNMR) αντί του FNMR.

4.4 *Σύγκριση Βιομετρικών Συστημάτων*

Η σύγκριση των βιομετρικών συστημάτων βασίζεται κυρίως στους δείκτες αξιολόγησης των βιομετρικών συστημάτων, οι οποίοι παρουσιάστηκαν στην παράγραφο 4.3 και στην παράγραφο 4.2.2.5.1. Προκειμένου να είναι αξιόπιστη η σύγκριση δεν αρκεί η σύγκριση ενός δείκτη. Έτσι, αν για παράδειγμα, οι κατασκευαστές των βιομετρικών συστημάτων παρέχουν μόνο την τιμή του Δείκτη Λανθασμένης Αποδοχής, δε μπορούν να εξαχθούν ασφαλή συμπεράσματα από τη σύγκριση των συστημάτων [61]. Αυτό οφείλεται στο γεγονός ότι πολύ χαμηλή τιμή του δείκτη αυτού θα συνοδεύεται από μη αποδεκτή τιμή του Δείκτη Λανθασμένης Απόρριψης.

Ακόμη όμως και στην περίπτωση που οι τιμές των δύο δεικτών δίνονται, η πληροφορία αυτή δεν είναι ασφαλής για τη σύγκριση των συστημάτων, καθώς οι τιμές αυτές εξαρτώνται από το επιλεγμένο κατώφλι. Έτσι, αν το κατώφλι είναι προσαρμόσιμο, δεν είναι εφικτή η απόφαση ότι ένα σύστημα με υψηλότερο Δείκτη Εσφαλμένης Αποδοχής και χαμηλότερο Δείκτη Εσφαλμένης Απόρριψης είναι καλύτερο από ένα σύστημα με χαμηλότερο Δείκτη Εσφαλμένης Αποδοχής και υψηλότερο Δείκτη Εσφαλμένης Απόρριψης.



Εικόνα 10 Συνήθεις επιλογές εφαρμογών για τα επίπεδα των FMR, FNMR

Προκειμένου η σύγκριση να είναι ανεξάρτητη από αυτό το κατώφλι χρησιμοποιείται ο Δείκτης Ίσου Σφάλματος. Έτσι, όσο μικρότερη είναι η τιμή αυτού του δείκτη τόσο μεγαλύτερη είναι η αποτελεσματικότητα του βιομετρικού συστήματος, καθώς αποτελεί και το σημείο στο οποίο το συνολικό σφάλμα (άθροισμα των δύο δεικτών) είναι μειωμένο. Παρ'όλα αυτά, ακόμη και η χρήση αυτού του δείκτη δουλεύει σωστά θεωρητικά, καθώς απαιτεί ένα απεριόριστο και αντιπροσωπευτικό δοκιμαστικό σύνολο, το οποίο και υπό συνθήκες πραγματικού κόσμου δεν είναι εφικτό. Για τον λόγο αυτό και προκειμένου να είναι συγκρίσιμα τα αποτελέσματα, είναι απαραίτητο ο δείκτης αυτός να υπολογίζεται για τα βιομετρικά συστήματα με βάση τα ίδια δοκιμαστικά δεδομένα χρησιμοποιώντας το ίδιο πρωτόκολλο δοκιμών. Προκειμένου να ξεπεραστεί αυτός ο περιορισμός σημαντική προσπάθεια λαμβάνει χώρα για τη συγκέντρωση μεγάλων και ευρέως διαθέσιμων συνόλων δοκιμαστικών δεδομένων όπως είναι οι βάσεις δεδομένων FERET (Facial Recognition Technology) [62] και XM2VTS (Extended Multimodal Verification for Teleservices and Security applications Database) [63] για αξιολόγηση συστημάτων αναγνώρισης προσώπου. Παρ' όλα αυτά περιορίζονται σε μονοτροπικά βιομετρικά συστήματα, καθιστώντας έτσι ιδιαιτέρως δύσκολη τη σύγκριση πολυτροπικών βιομετρικών συστημάτων.

5 *Βιομετρικά Συστήματα : Ασφάλεια και Ιδιωτικότητα*

Στο κεφάλαιο αυτό αναλύονται οι έννοιες της ασφάλειας και της ιδιωτικότητας και παρουσιάζεται η μελέτη μας σχετικά με τα κύρια τρωτά σημεία των βιομετρικών δεδομένων και οι παραβιάσεις στην ιδιωτικότητα των ατόμων ακολουθούμενη από μια καινοτόμο αντιστοίχηση αυτών. Επιπροσθέτως, περιγράφεται η δυνατότητα των βιομετρικών συστημάτων να αποτελέσουν τεχνολογίες διασφάλισης της ιδιωτικότητας.

5.1 *Οι έννοιες της ασφάλειας και της ιδιωτικότητας*

5.1.1 Η έννοια της ασφάλειας

Η έννοια της ασφάλειας αφορά στο βαθμό προστασίας έναντι κινδύνου, απώλειας και εγκλήματος. Τα άτομα ή οι πράξεις που παραβιάζουν τους όρους προστασίας είναι υπεύθυνα για ένα ρήγμα στην ασφάλεια [99]. Σύμφωνα με το Ινστιτούτο για Ασφάλεια και Ανοικτές Μεθοδολογίες (Institute for Security and Open Methodologies - ISECOM) η ασφάλεια είναι μια μορφή προστασίας ο διαχωρισμός της οποίας καθορίζεται από τα πλεονεκτήματα και την

απειλή. Προκειμένου να είναι ασφαλής, είτε το πλεονέκτημα αφαιρείται από την απειλή είτε η απειλή αφαιρείται από το πλεονέκτημα.

Συγκεκριμένα στον τεχνολογικό χώρο και κυρίως στον χώρο των πληροφοριακών συστημάτων, η ασφάλεια [100] αφορά στην προστασία της πληροφορίας και των πληροφοριακών συστημάτων από μη εξουσιοδοτημένη πρόσβαση, χρήση, διατάραξη, αποκάλυψη, τροποποίηση ή καταστροφή με στόχο την:

- Ακεραιότητα, η οποία αφορά στην προστασία έναντι ακατάλληλης τροποποίησης ή καταστροφής της πληροφορίας και περιλαμβάνει τη διασφάλιση της αυθεντικότητας της πληροφορίας,
- Εμπιστευτικότητα, η οποία αφορά στη διαφύλαξη των εξουσιοδοτημένων περιορισμών στην πρόσβαση και στην αποκάλυψη, συμπεριλαμβανομένων μέσων για την προστασία της προσωπικής ιδιωτικότητας και ιδιωτικής πληροφορίας
- Διαθεσιμότητα, η οποία αφορά στη διασφάλιση της έγκαιρης και αξιόπιστης πρόσβασης σε και χρήσης πληροφορίας

5.1.2 Η έννοια της ιδιωτικότητας

Η ιδιωτικότητα αποτελεί μια περίπλοκη και υποκειμενική έννοια με διαφορετική σημασία ανά άνθρωπο, η οποία εξαρτάται από την περίπτωση στην οποία χρησιμοποιείται. Ο Solove παρουσίασε μια ταξινόμηση της ιδιωτικότητας [64] από νομικής απόψεως, στην οποία ορίζονται 16 διαφορετικά είδη παραβίασης της ιδιωτικότητας. Αυτές οι παραβιάσεις της ιδιωτικότητας ταξινομούνται σε τέσσερεις κύριες κατηγορίες :

1. Συλλογή πληροφορίας: παρακολούθηση και ανάκριση

2. Επεξεργασία πληροφορίας: συγκέντρωση, αναγνώριση, ανασφάλεια, δευτερεύουσα χρήση και αποκλεισμός
3. Διάδοση πληροφορίας: παραβίαση εμπιστευτικότητας, αποκάλυψη, έκθεση, αυξημένη προσβασιμότητα, εκβιασμός, οικειοποίηση και παραποίηση
4. Εισβολή: παράνομη παρέμβαση και ανάμειξη στις αποφάσεις

Σύμφωνα με τον Clarke [65], η ιδιωτικότητα είναι το ενδιαφέρον που έχουν οι άνθρωποι στο να διατηρούν έναν «προσωπικό χώρο», ελεύθεροι από παρεμβάσεις από τρίτους είτε ανθρώπους είτε οργανισμούς. Στη μελέτη του τονίζει ότι η έννοια της ιδιωτικότητας δεν είναι μονοδιάστατη, αλλά αφορά στην:

- *Ιδιωτικότητα του ατόμου*, η οποία αναφέρεται συχνά και ως «σωματική ιδιωτικότητα» (‘bodily privacy’). Ο όρος αυτός αφορά στην ακεραιότητα του ανθρώπινου σώματος. Θέματα σχετικά αφορούν στην υποχρεωτική ανοσοποίηση, στη μετάγγιση αίματος δίχως συναίνεση, στην υποχρεωτική παροχή δειγμάτων σωματικών υγρών και σωματικού ιστού και στην υποχρεωτική στειροποίηση
- *Ιδιωτικότητα της προσωπικής συμπεριφοράς*, η οποία σχετίζεται με όλες τις όψεις της συμπεριφοράς, αλλά ιδιαιτέρως με ευαίσθητα θέματα, όπως είναι οι σεξουαλικές προτιμήσεις και οι συνήθειες, οι πολιτικές δραστηριότητες και οι θρησκευτικές πεποιθήσεις και δράσεις, τόσο σε ιδιωτικά όσο και σε δημόσια μέρη. Συχνώς αναφέρεται και ως «ιδιωτικότητα των μέσων» (‘media privacy’).
- *Ιδιωτικότητα της προσωπικής επικοινωνίας,* η οποία σχετίζεται με την επιθυμία των ανθρώπων να είναι σε θέση να επικοινωνούν μεταξύ τους χρησιμοποιώντας διάφορα μέσα, χωρίς συνήθη παρακολούθηση της επικοινωνίας τους με άλλα άτομα ή οργανισμούς. Είναι γνωστή και ως «ιδιωτικότητα υποκλοπής» (‘interception privacy’).

- *Ιδιωτικότητα προσωπικών δεδομένων*, η οποία αφορά στη μη διάθεση των προσωπικών δεδομένων των ανθρώπων σε τρίτους είτε ανθρώπους είτε οργανισμούς. Ακόμη και αν τα δεδομένα αυτά είναι στην κατοχή ενός άλλου, το άτομο θα πρέπει να ασκεί έλεγχο ουσιώδους βαθμού στα δεδομένα του και στη χρήση τους. Συχνά απαντάται και ως «ιδιωτικότητα δεδομένων» (‘data privacy’) ή «ιδιωτικότητα πληροφορίας» (‘information privacy’).

Οι λόγοι για τους οποίους η ιδιωτικότητα αποτελεί μια πολύ σημαντική παράμετρο και έκφανση της ανθρώπινης ζωής είναι ποικίλοι και σύμφωνα με τον Clarke μπορούν να κατηγοριοποιηθούν ως εξής:

- *Ψυχολογικά*: οι άνθρωποι έχουν την ανάγκη του ιδιωτικού χώρου. Αυτό αφορά τόσο σε δημόσιους χώρους όσο και κεκλεισμένων των θυρών. Η ανάγκη αυτή περιλαμβάνει τη δυνατότητα να εποπτεύουν τον χώρο, να κρίνουν αν οι άνθρωποι στην εγγύτητά τους αποτελούν απειλή, να κάνουν δραστηριότητες οι οποίες δυνητικά μπορεί να τους φέρουν σε αμηχανία
- *Κοινωνικά*: οι άνθρωποι έχουν ανάγκη να νιώθουν ελεύθεροι να συμπεριφέρονται και να συναναστρέφονται με άλλους, υποκείμενους σε ένα ευρύ φάσμα ηθικών αξιών, αλλά χωρίς τη διαρκή απειλή ότι παρακολουθούνται.
- *Οικονομικά*: οι άνθρωποι πρέπει να είναι ελεύθεροι να καινοτομήσουν.
- *Πολιτικά:* οι άνθρωποι έχουν ανάγκη να είναι ελεύθεροι να σκεφτούν και να διαφωνούν και να πράττουν.

Έτσι ο Clarke καταλήγει ότι η προστασία της ιδιωτικότητας είναι η διαδικασία εύρεσης των κατάλληλων ισορροπιών μεταξύ της ιδιωτικότητας και πολλαπλών ανταγωνιστικών συμφερόντων.

Τεχνολογικά μέτρα τα οποία αποσκοπούν στην προστασία από περιπτώσεις καταπάτησης της ιδιωτικότητας εστιάζουν κυρίως στο να αποτρέπουν τη μη αποσκοπούμενη διαρροή πληροφορίας. Έτσι τα τεχνολογικά συστήματα μπορούν κυρίως να προσφέρουν προστασία έναντι των ακολούθων απειλών έναντι της ιδιωτικότητας [66]:

- Παρακολούθηση: λαμβάνοντας υπ'όψιν τρίτους ικανούς να παρακολουθούν ηλεκτρονικές συναλλαγές, οι τεχνολογίες βελτίωσης της ιδιωτικότητας (Privacy-Enhancing Technologies - PETs) στοχεύουν στη μείωση του κινδύνου παρακολούθησης με την απόκρυψη πληροφορίας σχετικής με το περιεχόμενο και τις συνθήκες των ηλεκτρονικών συναλλαγών από τρίτους. Όταν οι χρήστες είναι σε θέση να κρατούν εμπιστευτικά τα περιεχόμενα των συναλλαγών τους και να δρουν ανωνύμως, προστατεύουν τους εαυτούς τους από απειλές παρακολούθησης.
- Ανάκριση: η τεχνική ιδιότητα η οποία προστατεύει έναν χρήστη από το να αναγκάζεται να αποκαλύψει πληροφορία καλείται αληθοφανής δυνατότητα άρνησης. Συστήματα τα οποία την παρέχουν καθιστούν αδύνατο για έναν τρίτο να αποδείξει ότι ο χρήστης αποκρύπτει πληροφορία
- Συγκέντρωση: η ιδιότητα η οποία αποτρέπει τη συγκέντρωση πληροφορίας η οποία σχετίζει δύο κομμάτια πληροφορίας ή με ένα συγκεκριμένο υποκείμενο καλείται μη συνδεσιμότητα
- Αναγνώριση: η αναγνώριση συνδέει δεδομένα με άτομα. Η ανωνυμία, η μη συνδεσιμότητα και η εμπιστευτικότητα αποτρέπουν την αποκάλυψη της σύνδεσης αυτής.

Προκειμένου να διατηρηθεί η ιδιωτικότητα σε ηλεκτρονικές εφαρμογές, η πληροφορία πρέπει να μην είναι διαθέσιμη σε μη εξουσιοδοτημένους τρίτους οι οποίοι προσπαθούν να

αναγνωρίσουν, να συνθέσουν προφίλ ή να συνδέσουν υποκείμενα με πράξεις, γνωρίσματα ή άλλα υποκείμενα.

5.2 *Τα τρωτά σημεία των βιομετρικών συστημάτων*

Εφαρμογές οι οποίες προσπαθούν να συμπεριλάβουν ένα βιομετρικό στοιχείο στη διαδικασία της ταυτοποίησης, επιβεβαίωσης ταυτότητας ή σύνθεσης προφίλ είναι ιδιαιτέρως επιρρεπής σε κινδύνους ασφάλειας – πολύ περισσότερο από τις απλές διαδικασίες, όπως είναι οι έξυπνες κάρτες ή οι κωδικοί. Με τις κυβερνήσεις να σχεδιάζουν την υιοθέτηση των βιομετρικών γνωρισμάτων στις διαδικασίες αυτές, οι βιομετρικές τεχνολογίες έχουν αποκτήσει στρατηγική αξία. Αυτή η στρατηγική τους αξία προέρχεται όχι μόνο από τα μακροπρόθεσμα κυβερνητικά σχέδια, αλλά και από τη μοναδική και κρίσιμη λειτουργία που τα ίδια ενσωματώνουν [60].

Μερικές κυβερνήσεις μάλιστα εκφράζουν τη μακροπρόθεσμη δέσμευσή τους στις βιομετρικές τεχνολογίες και προχωρούν πέραν από τις δικές τους διαδικασίες αναγνώρισης, προτείνοντας τη χρήση των δικών τους βάσεων δεδομένων από διάφορους επιχειρηματικούς χώρους, όπως είναι οι χρηματοπιστωτικοί οργανισμοί. Έτσι, για παράδειγμα, η Βρετανική κυβέρνηση οραματίζεται ότι οι επιχειρηματικοί κύκλοι θα είναι σε θέση να πιστοποιούν την ταυτότητα των πελατών τους μέσω της σύνδεσης με τις κυβερνητικές βάσεις δεδομένων πριν επιτρέψουν σε αυτούς να χρησιμοποιήσουν τις δικές τους υπηρεσίες [67].

Τέτοιες εφαρμογές και οι πολύπλοκες υποδομές οι οποίες σταδιακά υποστηρίζονται από βιομετρικές τεχνολογίες τόσο σε εθνικά όσο και σε διεθνή επίπεδα οδηγούν σε πραγματικά στρατηγικά πληροφοριακά συστήματα και αναδεικνύουν τη σπουδαιότητα των βιομετρικών τεχνολογιών σε κάτι πολύ περισσότερο από ένα ακόμη απλό τεχνολογικό εύρημα.

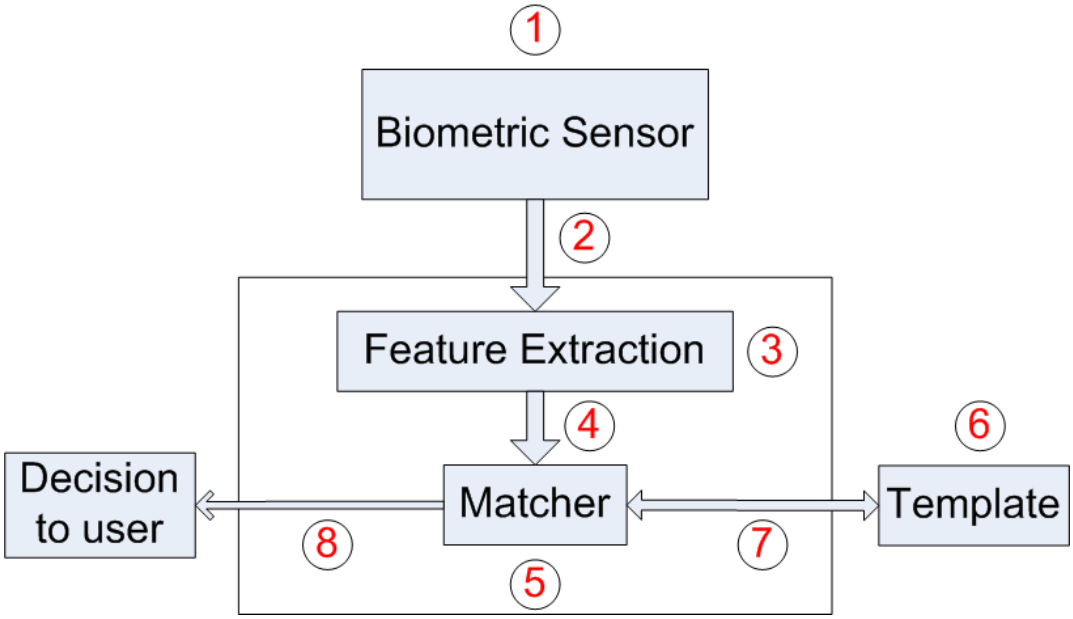
Λαμβάνοντας υπ'οψιν την ξεπροβάλλουσα σημασία αυτών των τεχνολογικών εφαρμογών οι οποίες στοχεύουν στο θέμα της ταυτοποίησης με την ενσωμάτωση βιομετρικών στοιχείων, είναι εξαιρετικά σημαντικό να μελετηθούν οι κύριες όψεις της ασφάλειας αυτών. Δεδομένου ότι οι βιομετρικές τεχνολογίες έχουν σημαντικές προεκτάσεις οι οποίες δε θα πρέπει να περιορίζονται στην αυστηρή τεχνολογική ανάλυσή τους, προχωρήσαμε σε μελέτη των θεμάτων ασφαλείας στις βιομετρικές εφαρμογές λαμβάνοντας ιδιαιτέρως υπ'όψιν το πεδίο εφαρμογής τους και διευρύνοντας σημαντικά τους παράγοντες τους οποίους συμπεριλάβαμε στη μελέτη μας.

Σε αυτό το σημείο θεωρούμε ιδιαιτέρως σημαντικό να τονίσουμε ότι ο όρος Ασφάλεια Πληροφοριακών Συστημάτων είναι αρκετά πιο ευρύς. Πέραν των αμιγώς τεχνικών του όψεων, υφίστανται επίσης επίσημες πλευρές οι οποίες σχετίζονται με αυτόν όπως διαδικασίες και πολιτικές, αλλά και ανεπίσημες και δυσδιάκριτες πλευρές οι οποίες μπορεί να επηρεάζουν την ασφάλεια μιας βιομετρικής υλοποίησης, όπως είναι οι πολιτιστικές και ηθικές νόρμες. Το τελευταίο απαιτεί επιπρόσθετες αρχές να λαμβάνονται υπ'όψιν για τη διαχείρηση της ασφάλειας των πληροφοριακών συστημάτων όσον αφορά στις βιομετρικές υλοποιήσεις, όπως είναι η ευθύνη, η εμπιστοσύνη και η ηθική οι οποίες θα πρέπει να είναι παρούσες για τη βελτίωση της ασφάλειας είτε εντός κυβερνήσεων είτε εντός οργανισμών [68].

Στις επόμενες παραγράφους η ανάλυσή μας εστιάζει στα τρωτά σημεία των βιομετρικών συστημάτων στα πλαίσια ευρύτερων υποδομών και όχι στις διαφορετικές βιομετρικές τεχνικές οι οποίες χρησιμοποιούνται, όπως είναι η ανάλυση δακτυλικών αποτυπωμάτων ή η αναγνώριση φωνής.

Μια από τις πιο συνήθεις παρεξηγήσεις η οποία σχετίζεται με τη βιομετρική ασφάλεια εντοπίζει τα περισσότερα ευάλωτα σημεία των βιομετρικών συστημάτων ως προς την ασφάλεια στη συσκευή ανάγνωσης των βιομετρικών χαρακτηριστικών (biometric reader device) ή στο

πρότυπο αναφοράς. Παρ'όλο που αυτά τα σημεία είναι τα πιο προφανή σημεία επίθεσης ενός βιομετρικού συστήματος, τα ζητήματα ασφάλειας είναι πολύ πιο πολύπλοκα με τους ερευνητές να παρουσιάζουν οκτώ τρωτά προς επίθεση σημεία [69], τα οποία και συνοψίζονται στην ακόλουθη εικόνα.



Εικόνα 11 Τα 8 τρωτά σημεία των βιομετρικών συστημάτων

Καθώς η αλληλεπίδραση χρήστη-συστήματος λαμβάνει χώρα μέσω ενός βιομετρικού αισθητήρα, όπως ήδη αναφέρθηκε, η συσκευή ανάγνωσης των βιομετρικών χαρακτηριστικών αποτελεί το πρώτο προφανές σημείο επίθεσης (point of attack). Συστήματα τα οποία μιμούνται βιομετρικά χαρακτηριστικά έχουν χρησιμοποιηθεί για τον σκοπό αυτόν, ενώ ακόμη και πληροφορίες σχετικές με τη συλλογή βιομετρικών χαρακτηριστικών μέσω απλών διαδικασιών απαντώνται στο Διαδίκτυο, όπως είναι η απομίμηση δακτυλικών αποτυπωμάτων [70]. Η επιτυχία των επιθέσεων με χρήση προσθετικών δακτύλων είναι υψίστης σημασίας από πλευράς ασφαλείας: ερευνητές από την Ιαπωνία έδειξαν δείκτη επιτυχίας για επιθέσεις με χρήση τέτοιων συσκευών από 67 ως και 100% [71]. Ο κύριος παράγοντας ο οποίος καθορίζει την επιτυχία των

επιθέσεων αποδείχθηκε ότι είναι η ποιότητα του αρχικού αποτυπώματος. Όπως θα δούμε και στις επόμενες ενότητες προκειμένου να περιοριστούν τέτοια συμβάντα είναι δυνατό τα βιομετρικά συστήματα να εμπλουτιστούν με μεθόδους ανίχνευσης ζωντάνιας (aliveness detection), όμως το κόστος αυτών αποτελεί περιοριστικό παράγοντα ενώ πολλά από τα εμπορικά συστήματα που ανιχνεύουν αν το υποκείμενο είναι εν ζωή μπορούν να ξεγελαστούν αρκετά εύκολα [72].

Πέραν όμως των ευθέων επιθέσεων στην ίδια τη συσκευή ανάγνωσης των βιομετρικών χαρακτηριστικών, υπάρχουν ακόμη επτά κύρια σημεία επίθεσης στα βιομετρικά συστήματα. Καθώς η βιομετρική συσκευή λαμβάνει πληροφορία από το παρεχόμενο βιομετρικό χαρακτηριστικό και το μετατρέπει σε δεδομένα για να το αποστείλει στο επόμενο δομικό στοιχείο του συστήματος, είναι δυνατό να γίνει παρεμβολή δεδομένων (δεύτερο σημείο επίθεσης). Στο τρίτο σημείο επίθεσης, παρέμβαση στο υποσύστημα εξαγωγής χαρακτηριστικών μπορεί να παράγει διαφορετικές τιμές από αυτές που θα προέκυπταν από την επεξεργασία της εξόδου του αισθητήρα. Στο τέταρτο σημείο επίθεσης, μη εξουσιοδοτημένη πρόσβαση μπορεί να επιτευχθεί σε ένα σύστημα με την αντικατάσταση των υπολογισθέντων τιμών από το υποσύστημα εξαγωγής βιομετρικών χαρακτηριστικών με άλλες γνωστές, ενώ στο πέμπτο σημείο επίθεσης το υποσύστημα σύγκρισης μπορεί να παράγει ψευδές υψηλό ή χαμηλό match score αναλόγως με τις προθέσεις των επιτιθέμενων κι έτσι να προσομοιωθεί ένα λανθασμένως θετικό ή αρνητικό αντιστοίχως.

Τα τελευταία τρία σημεία επίθεσης ενός βιομετρικού συστήματος παρουσιάζουν ιδιαίτερες προκλήσεις ασφάλειας. Παρά το γεγονός ότι το πρότυπο αναφοράς αποτελεί ένα προφανές και σύνηθες σημείο επίθεσης (στην περίπτωσή μας είναι το έκτο), εντούτοις εμπεριέχει τεράστιες προκλήσεις και είναι δυνατό να επιφέρει σημαντικά προβλήματα. Έτσι, μη εξουσιοδοτημένη

πρόσβαση (ή ακατάλληλη χρήση νόμιμης πρόσβασης από εργαζόμενους) μπορεί να έχει σημαντικές και πολύπλοκες επιπτώσεις, καθώς θα είναι δυνατή η δημιουργία, τροποποίηση ή διαγραφή προτύπων αναφοράς και επομένως ταυτοτήτων.

Μια πιο διευρυμένη όψη αυτού του σημείου επίθεσης αποτελεί η σύνδεση του καταχωρημένου προτύπου αναφοράς με ένα άλλο άτομο από αυτό στο οποίο ανήκει, το οποίο όμως σε μεγάλο βαθμό ξεπερνά τα τεχνολογικά όρια και κυρίως αφορά στα βήματα τα οποία λαμβάνουν χώρα αρχικά κατά την καταχώρηση των βιομετρικών δεδομένων αναφοράς στο σύστημα προκειμένου να πιστοποιηθεί η ταυτότητα του υποκειμένου το οποίο τα παρέχει. Έτσι, αν είτε τα καταχωρημένα βιομετρικά δεδομένα είτε τα προσωπικά δεδομένα τα οποία παρέχονται ανήκουν σε άλλο άτομο, τότε η σύνδεσή τους και η προκύπτουσα χρήση από τη μεριά του ατόμου αποτελούν σημαντικό ρήγμα στην ασφάλεια του συστήματος.

Το έβδομο σημείο επίθεσης αφορά στην υποκλοπή της μετάδοσης από τη βάση προτύπων αναφοράς προς το υποσύστημα σύγκρισης και στην μετατροπή του σε λανθασμένο θετικό ή λανθασμένο αρνητικό, ενώ το όγδοο σημείο επίθεσης αφορά σε επίθεση στο τελικό σημείο απόφασης του συστήματος, το οποίο στην περίπτωση επιβεβαίωσης ταχύτητας είναι δυαδικό (Ναι ή Όχι). Ειδικά στην περίπτωση προσβολής του όγδοου σημείου επίθεσης, τα βιομετρικά δεδομένα τα οποία έχουν υποβληθεί μέσω της συσκευής καθίστανται άσχετα. Ο επιτιθέμενος μπορεί να επιλέξει να αρνείται ή να προσφέρει την πρόσβαση σε οποιονδήποτε και γενικότερα να μεταβάλει το δικαίωμα πρόσβασης κατά την επιθυμία του.

Δεδομένου ότι, όπως αναφέραμε και προηγουμένως, η ασφάλεια των βιομετρικών συστημάτων αποτελεί ένα κομμάτι της ασφάλειας των πληροφοριακών συστημάτων, η ασφάλεια των υποδομών οι οποίες ενσωματώνουν βιομετρικές τεχνολογίες επηρεάζονται σημαντικά από κοινωνικούς και οικονομικούς παράγοντες, όπως και η διασύνδεσή τους. Ενώ η

τεχνολογία αυτή καθαυτή παραμένει σημαντική, το πλαίσιο εντός του οποίου η τεχνολογία λειτουργεί εντείνει υπάρχοντα και φέρνει ποικίλα νέα τρωτά σημεία στην ασφάλεια.

Στο γενικότερο πεδίο της ασφάλειας, παρά το γεγονός ότι οι κυβερνήσεις και οι οργανισμοί λαμβάνουν διαρκώς νέα επιπρόσθετα μέτρα για να διασφαλίσουν την ασφάλεια στα δεδομένα τους τα τελευταία χρόνια, τα ρήγματα στην ασφάλεια έχουν αυξηθεί. Σε έναν κόσμο ο οποίος γίνεται όλο και πιο εξαρτώμενος από την τεχνολογία, η ταμπέλα του «μη ασφαλούς» μπορεί να αποδειχθεί καταστροφικός για έναν οργανισμό [73]. Μόνο στο Ηνωμένο Βασίλειο, το ηλεκτρονικό έγκλημα κόστισε το 2005 2.4 δις αγγλικές λίρες με ένα τρομακτικό 90% των εταιρειών να έχουν υποστεί ηλεκτρονική «διάρρηξη» και μόνο το 1% αυτών να μην έχει οδηγήσει σε κλοπή δεδομένων [74]. Σύμφωνα με τη μελέτη αυτή οι επιθέσεις γίνονται όλο και πιο εκλεπτυσμένες και πιο περίτεχνες και παρατηρείται ένας ανερχόμενος επαγγελματισμός. Παράλληλα, ένας τέτοιος ανερχόμενος επαγγελματισμός εκμεταλλεύεται ολοένα και περισσότερο μεθόδους «κοινωνικής μηχανικής» προκειμένου να επιτεθεί στην ασφάλεια ποικίλων συστημάτων [75].

Τα παραπάνω σχόλια απεικονίζουν το πλαίσιο ασφαλείας το οποίο οι βιομετρικές τεχνολογίες καλούνται να καλύψουν. Με τις βιομετρικές τεχνολογίες και τα προηγούμενα σχόλια κατά νου, προκύπτει ένα επιπρόσθετο θέμα ασφάλειας το οποίο πρέπει να επιλυθεί. Ασχέτως με την αποδοτικότητα της τεχνολογίας αυτής καθ'εαυτής, οι μέθοδοι κοινωνικής μηχανικής για την υπονόμευση της ασφάλειας των βιομετρικών υλοποιήσεων θα πρέπει να τονιστούν και οι επιπτώσεις τους να συνυπολογιστούν και να διαχειριστούν με σύνεση. Έτσι, αν είναι δυνατή η δωροδοκία ενός υπαλλήλου ο οποίος έχει πρόσβαση στα βιομετρικά δεδομένα (και/ή στα πρότυπα αναφοράς), τότε οι τεχνικές πλευρές χάνουν την αξία τους. Η διαφθορά, πάντα παρούσα με τη μία ή την άλλη μορφή, θα βρει τον τρόπο να υπονομεύσει την ασφάλεια των

βιομετρικών συστημάτων. Για τον λόγο αυτόν είναι ύψιστης σημασίας οι αρχές της διαχείρησης πληροφορίας οι οποίες σχετίζονται με τις βιομετρικές υλοποιήσεις να υιοθετούνται για τον σκοπό αυτοόν και οι διαδικασίες να τοποθετούνται ώστε να διασφαλίζουν την αποτελεσματική διαχείρηση και τον χειρισμό γεγονότων σε περιπτώσεις κατά τις οποίες λαμβάνουν χώρα ρήγματα ασφάλειας.

Η υιοθέτηση διεθνών προτύπων όπως είναι το ISO17799 (το οποίο τώρα έχει αναχθεί σε 27001) τα οποία ασχολούνται με την ασφάλεια των πληροφοριακών συστημάτων γίνεται αναγκαιότητα δεδομένης της φύσης των βιομετρικών δεδομένων. Έτσι οι κυβερνήσεις και οι επιχειρήσεις θα πρέπει να αναζητούν πιστοποιήσεις έναντι σχετικών διεθνών προτύπων ασφάλειας. Υποστήριξη από τις υψηλές βαθμίδες διαχείρισης, κατανόηση των κινδύνων και των πλεονεκτημάτων της πληροφορίας, διάδοση μιας σαφούς πολιτικής ασφάλειας στους εργαζόμενους, όπως και αναθεώρηση και αξιολόγηση της απόδοσης αποτελούν κρίσιμα στοιχεία ανάμεσα στις πολλές πλευρές επιτυχίας των προτύπων ασφαλείας [76].

Θα πρέπει να τονιστεί ότι ενώ αυτή η ενότητα παρουσίασε τα ευάλωτα σημεία των βιομετρικών συστημάτων όσον αφορά στην ασφάλειά τους, δε θα πρέπει όπως έχουμε δει να παραμερήσουμε τη βασική τους λειτουργία, τη διασφάλιση της ασφάλειας. Τα βιομετρικά συστήματα εκμεταλλευόμενα τις εγγενείς ιδιότητες των βιομετρικών χαρακτηριστικών στοχεύουν και υπόσχονται τη βελτίωση της ασφάλειας μέσω πιο αξιόπιστης ταυτοποίησης και πιστοποίησης ταυτότητας.

5.3 *Καινοτόμος διασύνδεση θέματων ιδιωτικότητας και ασφάλειας στα βιομετρικά συστήματα*

Όπως περιγράψαμε και στις προηγούμενες παραγράφους, τα υπάρχοντα βιομετρικά συστήματα περιλαμβάνουν πιθανούς κινδύνους ασφαλείας οι οποίοι με τη σειρά τους μπορούν να οδηγήσουν σε πιθανούς κινδύνους για την ιδιωτικότητα. Αυτές οι απειλές προς την ιδιωτικότητα εντείνονται από τις τεχνολογικές αδυναμίες των βιομετρικών συστημάτων, καθώς κανένα βιομετρικό σύστημα δε μπορεί να προσφέρει 100% επιτυχία στη σωστή ταυτοποίηση ή επιβεβαίωση ταυτότητας των ατόμων. Επίσης, όπως είδαμε η προσπάθεια αύξησης των αληθώς θετικών συνοδεύεται από ταυτόχρονη αύξηση των λανθασμένως αρνητικών.

Το Electronic Privacy Information Centre (EPIC) [78] αναγνωρίζει τέσσερεις βασικές έννοιες της ιδιωτικότητας [79] (παρόμοιες με αυτές που είδαμε στην παράγραφο 5.1.2): α) *τη σωματική ιδιωτικότητα*, β) *τη χωρική ιδιωτικότητα*, γ) *την ιδιωτικότητα της πληροφορίας* και δ) *την πληροφοριακή ιδιωτικότητα*. Η πρώτη αναφέρεται στην προστασία του σώματος έναντι επεμβατικών διαδικασιών, ενώ η δεύτερη θέτει όρια στην εισβολή σε οικειακά και άλλα περιβάλλοντα. Η ιδιωτικότητα της πληροφορίας περιλαμβάνει κανόνες για τον χειρισμό προσωπικών δεδομένων και την ιδιωτικότητα των επικοινωνιών με τη μορφή αλληλογραφίας, τηλεφώνου και άλλων μορφών επικοινωνίας. Η *πληροφοριακή ιδιωτικότητα,* η οποία ενσωματώνει έναν πιο περιγραφικό ορισμό και αποτελεί την κύρια όψη της ιδιωτικότητας με την οποία σχετίζονται οι βιομετρικές τεχνολογίες, αποτελεί τη σύσταση κανόνων οι οποίοι κυβερνούν τη συλλογή και τον χειρισμό προσωπικών δεδομένων όπως είναι οι πληροφορίες πιστωτικών καρτών, ιατρικά και κυβερνητικά αρχεία, (γνωστό και ως «προστασία δεδομένων»)

με κάθε δευτερεύουσα χρήση της πληροφορίας αυτής να αποτελεί παραβίαση του δικαιώματος του ατόμου να την ελέγχει [77].

Όπως έχει ήδη αναφερθεί, τα βιομετρικά δεδομένα αποτελούν ιδιαιτέρως προσωπικά δεδομένα με τη μεγαλύτερη δύναμή τους αλλά και τη μεγαλύτερη απειλή τους προς την ιδιωτικότητα να προκύπτει από τη στενή τους σύνδεση με την ταυτότητα του κατόχου τους. Η σταδιακά αυξανόμενη χρήση των βιομετρικών τεχνολογιών σε διάφορα πεδία και το όραμα της εφαρμογής τους ως μοναδικά αναγνωριστικά σε μεγάλης κλίμακας – ακόμη και παγκόσμιες – εφαρμογές μπορεί να αποτρέψει του ερασιτέχνες κλέφτες. Όπως, η αυξανόμενη αξία των βιομετρικών δεδομένων ως μοναδικοί ταυτοποιητές οι οποίοι είναι δυνατό να προσφέρουν πρόσβαση σε ποικίλες εφαρμογές σε πολυάριθμα πεδία τροφοδοτεί μια πρόκληση προς αποφασισμένους «επαγγελματίες» να αποκτήσουν στην κατοχή τους τα δεδομένα αυτά προκειμένου να έχουν πρόσβαση σε αυτά, να τα τροποποιούν ή να τα προσπερνούν.

Ακολουθώντας την ανάλυση ασφάλειας η οποία παρουσιάστηκε στην Εικόνα 11, οι προσπάθειες ενός εισβολέα ο οποίος δεν έχει γνώσεις τεχνολογίας ή πρόσβαση στις βάσεις με τα πρότυπα αναφοράς θα εστιάσουν στο πρώτο σημείο επίθεσης μέσω της υποβολής ψεύτικων ή κλεμμένων βιομετρικών δεδομένων, μια βασική περίπτωση απάτης ταυτότητας. Σε υπάρχοντα συμβατικά συστήματα ταυτοποίησης ή επιβεβαίωσης ταυτότητας, όπως είναι τα συστήματα τα οποία βασίζονται σε πιστωτικές κάρτες ή κωδικούς, η αδυναμία σε επιθέσεις οι οποίες κυρίες στηρίζονται σε απάτη ταυτότητας αποτελούν ένα από τα κύρια ζητήματα [80]. Στις περιπτώσεις αυτές η απάτη ταυτότητας αναφέρεται στην κλοπή και χρήση ταυτοποιητών με στόχο κυρίως την απόκτηση των προνομίων των ατόμων προκειμένου να έχουν ειδικές άδειες, κρύβωντας μια προσωπική ταυτότητα ή διαπράττωντας οικονομική απάτη, ξέπλυμα χρήματος, ηλεκτρονικά εγκλήματα ή ακόμη και τρομοκρατικές δράσεις. Η υλοποίηση συστημάτων ταυτοποίησης ή

πιστοποίησης βασισμένων μερικώς ή πλήρως σε βιομετρικές τεχνολογίες, ή η ενσωμάτωση βιομετρικών τεχνολογιών σε υπάρχοντα συστήματα, στοχεύει στην εκμετάλλευση των δυνατοτήτων των βιομετρικών χαρακτηριστικών, δυνατότητες οι οποίες βασίζονται στη μοναδικότητα κάθε ατόμου και στη δυσκολία της εύκολης και ακριβούς αναπαραγωγής τους.

Παρ'όλα αυτά τόσο η σωματική όσο και η πληροφοριακή ιδιωτικότητα διακυβεύονται. Η πρώτη απειλείται στο πρώτο σημείο επίθεσης με τον εξαναγκασμό του ατόμου να εισαγάγει τα βιομετρικά του δεδομένα στο σύστημα ή ακόμη και με την κλοπή τους (π.χ. κοπή δακτύλου), η οποία κυρίως λαμβάνει χώρα από σχετικά ερασιτέχνες εγκληματίες, οι οποίοι προσπαθούν να επιτεθούν στο βιομετρικό σύστημα μέσω της άμεσης συλλογής της βιομετρικής πληροφορίας από τον κάτοχό της.

Όσον αφορά στην πληροφοριακή ιδιωτικότητα, τα κύρια σημεία επίθεσης τα οποία αποτελούν επέμβαση σε αυτήν είναι το πρώτο και το έκτο σημείο επίθεσης, μέσω της συλλογής ή της πλαστογράφησης των βιομετρικών ιχνών του ατόμου. Σε μια προσπάθειά μας να συγκρίνουμε τον αντίκτυπο των δύο αυτών σημείων επίθεσης στην πληροφοριακή ιδιωτικότητα, μια επίθεση στο πρότυπο αναφοράς (έκτο σημείο επίθεσης) των βιομετρικών δεδομένων του ατόμου – εισαγωγή, τροποποίηση, διαγραφή – θεωρείται πολύ πιο σοβαρή απειλή, καθώς επιτρέπει μια σειρά από λανθασμένα θετικά μέχρι τον εντοπισμό της. Αυτή η απειλή λαμβάνει ακόμη μεγαλύτερες διαστάσεις από το γεγονός ότι τα βιομετρικά χαρακτηριστικά αποτελούν μόνιμους ταυτοποιητές, σε αντίθεση με έναν κωδικό ή μια έξυπνη κάρτα, ενώ οι εναλλακτικές επιλογές είναι περιορισμένες (δύο μάτια, δέκα δάκτυλα των χεριών). Για τον λόγο αυτόν η *βιομετρική εξαπάτηση* (*biometrics spoofing*) αποτελεί τόσο κίνδυνο όσο και πρόκληση την οποία καλούνται τα βιομετρικά συστήματα να αντιμετωπίσουν, ιδιαιτέρως αφού τα βιομετρικά χαρακτηριστικά γενικώς εκτίθενται δημοσίως (εικόνες του προσώπου μπορούν εύκολα να ληφθούν, δακτυλικά

αποτυπώματα μπορούν να συλλεγούν από κάθε σημείο το οποίο ακουμπά το άτομο). Οι τεχνολογικές λύσεις για πολλά από τα παραπάνω θέματα είναι γνωστές ως τεχνικές αντι-εξαπάτησης (anti-spoofing).

Οι κύριες αρχές οι οποίες υποστηρίζουν την προστασία των δεδομένων περιλαμβάνουν την ελαχιστοποίηση των δεδομένων (data minimisation), τη θεμιτή επεξεργασία (lawful processing), τον έλεγχο του υποκειμένου των δεδομένων, τον περιορισμό της αποκάλυψης δεδομένων (disclosure limitation) και την προδιαγραφή σκοπού (purpose specification) [79]. Μια έντονη ανησυχία η οποία πλαισιώνει τις κυβερνητικές και εμπορικές εφαρμογές των βιομετρικών τεχνολογιών είναι η συστηματική συλλογή και χρήση των βιομετρικών δεδομένων η οποία μπορεί να καταλήξει να είναι περιττή και κυρίως μη εγκεκριμένη [81]. Στην περίπτωση ειδικά μη επεμβατικών βιομετρικών τεχνολογιών, η συλλογή των βιομετρικών δεδομένων μπορεί να λάβει χώρα ακόμη και χωρίς τη γνώση του ατόμου. Με τον όρο «μη επεμβατικές βιομετρικές τεχνολογίες» αναφερόμαστε σε βιομετρικές τεχνολογίες οι οποίες δεν ανιχνεύονται εύκολα από το υποκείμενο, κυρίως λόγω του ότι οι τεχνολογίες αυτές λειτουργούν χωρίς να απαιτούν αλληλεπίδραση με το άτομο, όπως είναι η αναγνώριση προσώπου μέσω κάμερας. Λαμβάνοντας υπ’όψιν το γεγονός ότι τέτοιες τεχνολογίες αποτελούν μέρος έξυπνων συστημάτων παρακολούθησης τα οποία εφαρμόζονται είτε σε εγκαταστάσεις δημόσιας πρόσβασης είτε σε ιδιωτικές περιοχές (όπως για παράδειγμα σε γήπεδα ή γραφεία) για λόγους ασφαλείας, ο φόβος ότι ένα ηλεκτρονικό ίχνος των κινήσεων κάθε ατόμου επεξεργάζεται και αποθηκεύεται είναι ρεαλιστικός με τη δύναμη των αρχών να αυξάνεται και τις κοινωνικές ελευθερίες να παραβιάζονται.

Καθώς η ανωνυμία θεωρείται η ικανότητα των ανθρώπων να ζουν και να εργάζονται δίχως να κάνουν γνωστές τις δραστηριότητές τους σε τρίτους, το δικαίωμα διατήρησης της ανωνυμίας

στην καθημερινή ζωή διακυβεύεται [82] [83]. Οι υποκλοπές οι οποίες εκμεταλλεύονται τις δυνατότητες των τεχνολογιών αναγνώρισης φωνής και παρακολούθησης μέσω κάμερας θεωρούνται από πολλούς ως μια εξαιρετικά επεμβατικής μορφής παρακολούθηση και μια διερευνητική τεχνική η οποία θα πρέπει να χρησιμοποιείται μόνο υπό ειδικές συνθήκες και όχι συχνά. Επιπροσθέτως, αυτό που είναι ιδιαιτέρως ανησυχητικό είναι το γεγονός ότι πολλές κάμερες οι οποίες λαμβάνουν εικόνες από πλατείες, γωνίες δρόμων ή λεωφόρους μπορούν να προσφέρουν πρόσβαση μέσω του Διαδικτύου.

Η υλοποίηση βιομετρικών διαβατηρίων επίσης προκαλεί σημαντικές ανησυχίες όσον αφορά στην προστασία των ταξιδιωτικών δεδομένων. Σε μια προσπάθειά μας να τονίσουμε αυτού του είδους τις απειλές στα πλαίσια του ερευνητικού προγράμματος FIDIS (Future of Identity in the Information Society) [84], οι ερευνητές συγγράψαμε και δημοσιεύσαμε τη «Διακύρηξη της Βουδαπέστης για τα Νέα Ταξιδιωτικά Έγγραφα» (“Budapest Declaration on Machine Readable Travel Documents (MRTDs)”) [85] η οποία μεταξύ των άλλων συμπεριελάμβανε συστάσεις προς κυβερνήσεις και βιομηχανίες. Αν και αυτό το ζήτημα δεν είναι τόσο ανησυχητικό όσο η διακύβευση των οικονομικών ή των ιατρικών δεδομένων ενός ατόμου, φόβοι έχουν εκφραστεί ότι οι ταξιδιώτες τώρα πια θεωρούνται και αντιμετωπίζονται ως πιθανοί τρομοκράτες και επομένως στόχοι παρακολούθησης [86]. Λαμβάνοντας υπ’όψιν ότι μερικά από τα συστήματα αυτά, στοχεύοντας σε μεγαλύτερα επίπεδα ασφάλειας μέσω της γνώσης καταστάσεων (για παράδειγμα αναγνώριση ανώμαλης συμπεριφοράς σε σχέση με τη συνήθη), δεν πραγματοποιούν μόνο ιχνηλάτηση ταυτότητας αλλά και αναγνώριση και ιχνηλάτηση δραστηριοτήτων [58] [59], και ότι το πλήθος των καμερών και άλλων συσκευών αυξάνεται διαρκώς, οι κίνδυνοι για την ιδιωτικότητα καθίστανται πολύ πιο μεγάλοι απ’ότι αναμενόταν. Ουσιαστικά, όλες οι προαναφερθείσες «παρενέργειες» στην ιδιωτικότητα αποτελούν διαφορετικές και υψίστης

σημασίας εκδηλώσεις του προβλήματος επαναπροσδιορισμού στόχου των δεδομένων (data re-purposing).

Μια σημαντική κοινωνική απειλή αφορά στη χρήση των βιομετρικών συστημάτων παρακολούθησης προκειμένου να παρακολουθούνται οι δραστηριότητες συγκεκριμένων ομάδων ατόμων. Ιδιαιτέρως μετά την 11η Σεπτεμβρίου, παρατηρείται μια έντονη ρυθμιστική πίεση, ενώ η ελευθερία της πληροφορίας, η ιδιωτικότητα και η ελευθερία του λόγου είναι υπό διαρκή εξονυχιστική έρευνα [87]. Πληροφορία η οποία κατηγοριοποιεί τα άτομα σε αυτές τις ομάδες μπορεί να είναι είτε το αποτέλεσμα αποκάλυψης πληροφοριών από εταιρείες ή κυβερνήσεις – παραβιάζοντας την αρχή προστασίας δεδομένων περί περιορισμού αποκάλυψης – ή μέσω τις περαιτέρω επεξεργασίας των βιομετρικών δεδομένων και της εξαγωγής soft biometrics. Όπως έχουμε δει και στην παράγραφο 3.3 αφορούν σε χαρακτηριστικά όπως το φύλο και η εθνικότητα. Εφαρμογές βιομετρικών συστημάτων τα οποία ενσωματώνουν soft biometrics περιλαμβάνουν από φιλτράρισμα των υποκειμένων των οποίων τα στοιχεία είναι καταχωρημένα στις βάσεις αναφοράς κατά την ταυτοποίηση μέχρι και εφαρμογές στατιστικής. Αν και η χρήση τους βελτιώνει την απόδοση των βιομετρικών συστημάτων τόσο σε ταχύτητα όσο και σε αξιοπιστία, εντούτοις το κύριο πεδίο εφαρμογής τους δεν είναι η ταυτοποίηση και η επιβεβαίωση ταυτότητας, όπως έχουμε δει. Φόβοι σχετικά με τη χρήση τέτοιων εφαρμογών στρέφονται γύρω από την παροχή ειδικών προνομίων σε συγκεκριμένες ομάδες ατόμων και την άρνηση εξυπηρέτησης σε άτομα άλλων ομάδων οδηγώντας έτσι σε διακρίσεις και ρατσισμό.

Ένα άλλο σημαντικό ζήτημα αφορά στα δεδομένα τα οποία αποθηκεύονται στο βιομετρικό σύστημα. Συγκεκριμένα, κατά τη φάση της καταχώρησης τα αποθηκευμένα δεδομένα μπορεί να είναι είτε τα ληφθέντα βιομετρικά δεδομένα αυτούσια (raw data), όπως είναι η εικόνα του προσώπου, είτε μόνο τα επεξεργασμένα δεδομένα αναφοράς (template). Επίσης το σύστημα

μπορεί να λειτουργεί σε πραγματικό χρόνο ή να προχωρά σε ταυτοποίηση των ατόμων σε κάποια μελλοντική στιγμή. Όσον αφορά στα δεδομένα τα οποία αποθηκεύονται κατά τη φάση της καταχώρησης, από τεχνικής άποψης είναι προτιμητέο να αποθηκεύονται τα βιομετρικά δεδομένα αυτούσια, καθώς αυτό προσφέρει στο σύστημα ευελιξία σε σχέση με την εκμετάλλευση του πλούτου των βιομετρικών δεδομένων για την ταυτοποίηση καθώς και δυνατότητα αναβάθμισης με εισαγωγή νέων, πιο αποτελεσματικών αλγόριθμων εξαγωγής χαρακτηριστικών, χωρίς να απαιτείται η επαναϋποβολή των βιομετρικών δεδομένων του ατόμου στο σύστημα από το ίδιο το άτομο.

Όμως, αυτός ο πλούτος της βιομετρικής πληροφορίας στα αρχικά μη επεξεργασμένα βιομετρικά δεδομένα μπορεί να παρέχει επιπρόσθετη πληροφορία για το ίδιο το άτομο πέραν της εξουσιοδοτημένης χρήσης του συστήματος και εν αγνοία του ατόμου. Χαρακτηριστικά παραδείγματα αυτής αποτελούν η κατάσταση της υγείας του και η εθνικότητά του. Τα παραπάνω ισχύουν και στις περιπτώσεις μεταγενέστερης επεξεργασίας των υποβληθέντων βιομετρικών δεδομένων κατά τη φάση της επερώτησης. Παράλληλα, μια επίθεση στο έκτο σημείο επίθεσης ενός βιομετρικού συστήματος (δηλαδή στη βάση των βιομετρικών δεδομένων) στην περίπτωση αυτή αποτελεί πολύ μεγαλύτερη απειλή από την περίπτωση που είναι αποθηκευμένα τα κωδικοποιημένα επεξεργασμένα βιομετρικά πρότυπα αναφοράς. Για τους λόγους αυτούς συστήνεται η μη διατήρηση των αρχικών αυτούσιων βιομετρικών δεδομένων μετά τη σύνθεση των βιομετρικών προτύπων αναφοράς [81].

Δεδομένου του αυστηρού προσωπικού χαρακτήρα των βιομετρικών χαρακτηριστικών, τα τελευταία έχουν την εγγενή δυνατότητα να αποτελέσουν παγκόσμιους ανθρώπινους ταυτοποιητές οι οποίοι θα συνδέουν μεταξύ τους δεδομένα για ένα άτομο σχετικά με την καθημερινή του ζωή, τις οικονομικές συναλλαγές, τα ταξίδια και άλλα ευαίσθητα δεδομένα,

όπως το ιατρικό του αρχείο, οι θρησκευτικές του πεποιθήσεις, η πολιτική του δράση, το ποινικό του μητρώο, συνθέτοντας με τον τρόπο αυτόν μια σχεδόν πλήρη εικόνα του ατόμου και των δραστηριοτήτων του. Ισοδυναμώντας με έναν συνδυασμό παραβιάσεων των αρχών της προστασίας δεδομένων περί περιορισμού αποκάλυψης, του ελέγχου των δεδομένων από το υποκείμενο και προσδιορισμού σκοπού, τέτοιου είδους σύνδεση πληροφοριών αποτελεί σημαντικό κίνδυνο για την ιδιωτικότητα.

Συνεπώς, αν και η σύνδεση δεδομένων μπορεί να οδηγήσει σε καλύτερη παροχή υπηρεσιών με βελτιωμένη εξατομίκευση, μπορεί επίσης να καταστρατηγήσει την ιδιωτικότητα του ατόμου, μέσω της συστηματικής σύνθεσης και χρήσης ενός αναλυτικού προφίλ του χωρίς τη συναίνεσή του. Χαρακτηριστικό παράδειγμα αυτού αποτελεί η πιθανότητα σύνδεσης των αρχείων πελατών μιας ασφαλιστικής εταιρείας με τα ιατρικά τους αρχεία. Οι περιπτώσεις αυτές εντάσσονται στο γενικότερο ζήτημα της μη εξουσιοδοτημένης χρήσης πληροφοριών. Έτσι, όταν οι καταναλωτές δίνουν τη συναίνεσή τους για τη χρήση ενός μοναδικού βιομετρικού τους χαρακτηριστικού για να πραγματοποιούν τις συναλλαγές τους με τα αυτόματα μηχανήματα αναλήψεων των τραπεζών ή για να έχουν πρόσβαση σε ένα υπολογιστικό κέντρο, υποθέτουν ότι τα βιομετρικά τους δεδομένα θα χρησιμοποιηθούν αποκλειστικά για τον συγκεκριμένο σκοπό. Ιδιαιτέρως στον ιδιωτικό τομέα στον οποίο οι περισσότερες αποφάσεις λαμβάνονται με γνώμονα το κέρδος, υπάρχει διάχυτος ο φόβος χρήσης των βιομετρικών τεχνολογιών και των δυνατοτήτων τους με παράλληλη αδιαφορία για το κόστος στην ιδιωτικότητα. Παράδειγμα αυτού αποτελεί η εξαγωγή ιατρικής πληροφορίας από περισυνελλεγμένα βιομετρικά δεδομένα – και κυρίως το DNA, τα δακτυλικά αποτυπώματα και την ίριδα – δίχως τη γνώση και τη συγκατάθεση του ατόμου και η χρήση αυτών ενάντια στα συμφέροντα του ατόμου, όπως είναι η διάκριση.

5.4 *Βιομετρικά συστήματα και προστασία ιδιωτικότητας*

Οι υπάρχουσες αδυναμίες των βιομετρικών συστημάτων όσον αφορά στην ασφάλεια και η πιθανή κατάχρηση των βιομετρικών δεδομένων συνδυασμένες με το διαρκώς εντονότερο εμπορικό και κυβερνητικό ενδιαφέρον στις τεχνολογίες αυτές προέρχεται, όπως έχουμε ήδη αναφέρει, από τις εγγενείς δυνατότητές τους για ταυτοποίηση οι οποίες υπόσχονται να ξεπεραστούν τα προβλήματα των συμβατικών συστημάτων ταυτοποίησης και επιβεβαίωσης ταυτότητας με τις ίδιες αυτές τις δυνατότητες να προκαλούν σημαντικές ανησυχίες για την ιδιωτικότητα του ατόμου. Όμως, λαμβάνοντας υπ'όψιν ότι υπόσχονται βελτίωση των επιπέδων ασφαλείας λόγω αυτών των δυνατοτήτων τους, οι βιομετρικές τεχνολογίες δεν αποτελούν μόνο απειλή για την ιδιωτικότητα αλλά μπορούν να αποτελέσουν και μια καλή ευκαιρία για τη διασφάλισή της.

Μια τεχνολογία χαρακτηρίζεται ως *privacy-enhancing* όταν προστατεύει την ιδιωτικότητα με το να αποτρέπει την περιττή ή μη εξουσιοδοτημένη αποκάλυψη και επεξεργασία δεδομένων, αλλά παράλληλα επιτρέπει την ομαλή λειτουργία του συστήματος [88]. Ένα βιομετρικό (υπο-) σύστημα σχεδιασμένο με προσανατολισμό βελτίωσης της ιδιωτικότητας και προσφέροντας τα απαιτούμενα επίπεδα ασφαλείας μπορεί να αποτελέσει *φύλακα της ιδιωτικότητας*. Η πιο προφανής χρήση των βιομετρικών τεχνολογιών ως τεχνολογία που προστατεύει την ιδιωτικότητα είναι ως μέσο ελέγχου της πρόσβασης στα προσωπικά δεδομένα των ατόμων μέσω ενός αυστηρού και αξιόπιστου συστήματος ελέγχου πρόσβασης. Τα υψηλά επίπεδα λανθασμένης αποδοχής και απόρριψης των υπαρχόντων συμβατικών συστημάτων ταυτοποίησης και επιβεβαίωσης ταυτότητας οδηγούν σε περιττές διπλοεγγραφές, συχνές περιπτώσεις κλοπής ταυτότητας και συνακόλουθη ενόχληση των πελατών και για τον λόγο αυτόν επιφέρουν σημαντικό κόστος στους οργανισμούς. Η ίδια η φύση των βιομετρικών συστημάτων είναι η

χρήση πληροφορίας του «ποιος είσαι» για σκοπούς πρόσβασης σε αντίθεση με τα συμβατικά συστήματα ταυτοποίησης και ελέγχου πρόσβασης τα οποία βασίζεται στο «τι έχεις» ή «τι γνωρίζεις» έχοντας το πλεονέκτημα ότι βασίζονται σε σύνολο χαρακτηριστικών τα οποία είναι στενά συνδεδεμένα με την ταυτότητα του ατόμου τα οποία δε μοιράζονται, μαντεύονται ή πλαστογραφούνται εύκολα. Αυτό ακριβώς το πλεονέκτημα σε συνδυασμό με το γεγονός ότι δεν απαιτείται η διατήρηση πολλαπλών μέσων ταυτοποίησης (όπως πιστωτικές κάρτες, κωδικούς, κλπ) [89].

Η έννοια της προστασίας της ιδιωτικότητας περιλαμβάνει την ικανότητα διόρθωσης ενός σφάλματος ή αποτροπής απάτης η οποία σχετίζεται με μια υποπτευόμενη κατάχρηση ή χρήση κατά αδόκιμο τρόπο προσωπικής πληροφορίας από τρίτους. Σύμφωνα με τον Neuman, ένα βιομετρικό χαρακτηριστικό (και συνακολούθως μια βιομετρική τεχνολογία) μπορεί να προστατέψει την ιδιωτικότητα, όπως όταν μια αρχή ψάχνει για κάτι εντός ενός συστήματος και η εξουσία της να το κάνει αυτό επίσης επιβεβαιώνεται μέσω της παροχής του ίδιου της του βιομετρικού γνωρίσματος [90]. Αλλά γενικώς, η πρόσβαση σε βάσεις δεδομένων και σε άλλου είδους αποθηκευτικά μέσα δεδομένων τα οποία περιέχουν ευαίσθητα προσωπικά δεδομένα όπως θρησκευτικά, οικονομικά, ιατρικά, ποινικά και άλλα, μπορεί να παρακολουθηθεί και να καταγραφεί με τη χρήση βιομετρικών τεχνολογιών πρόσβασης.

Παραδείγματα τέτοιων εφαρμογών είναι:

- Πρόσβαση σε προσωπική πληροφορία – ιατρική μεταξύ των άλλων – η οποία μπορεί να περιοριστεί σε ιατρικό προσωπικό το οποίο έχει δικαιώματα πρόσβασης μέσω βιομετρικά προστατευόμενων έξυπνων καρτών
- Πρόσβαση σε στοιχεία επικοινωνίας για έκτακτα περιστατικά και ειδική ιατρική πληροφορία μαθητών η οποία απαιτεί την παρουσία του μαθητή

- Πρόσβαση σε συγκεκριμένες εγκαταστάσεις και συσκευές ενός εργαστηρίου η οποία είναι εφικτή μόνο για προσωπικό του εργαστηρίου

Επιπροσθέτως, η χρήση των βιομετρικών τεχνολογιών ως τεχνολογίες σύνθεσης προφίλ [91] ή ως σύνδεσμος μεταξύ προσωπικών δεδομένων [60] μπορεί να αποτελέσει σημαντικό προσόν, δεδομένου ότι είναι υλοποιημένες και λειτουργούν υπό κατάλληλο νομικό καθεστώς – όταν λαμβάνουν χώρα έρευνες για τη διασφάλιση ότι ένα άτομο δεν έχει αρνητικό ιστορικό, ιδιαιτέρως σε περιπτώσεις παιδικής κακοποίησης ή βιασμών [89].

Για τον λόγο αυτόν, πέραν από την κατάρτιση ενός κατάλληλου νομοθετικού πλαισίου, πολλες τεχνολογικές προσπάθειες έχουν λάβει χώρα προκειμένουν να εξαληφθούν οι αδυναμίες των βιομετρικών συστημάτων οι οποίες απειλούν την ιδιωτικότητα των ατόμων και επισκιάζουν τις δυνατότητες των βιομετρικών συστημάτων για προστασία της ιδιωτικότητας. Θα πρέπει να τονιστεί ότι αυτές οι προσπάθειες στοχεύουν στο να προσφέρουν ένα ωφέλιμο επιπρόσθετο στρώμα στην όψη της ιδιωτικότητας των βιομετρικών συστημάτων και δεν παρέχουν μια πλήρη λύση.

Όσον αφορά έτσι στα ίδια τα βιομετρικά συστήματα η κεντρικοποιημένη αποθήκευση των βιομετρικών δεδομένων γενικώς αποθαρρύνεται, ενώ έννοιες όπως *ακυρώσιμα βιομετρικά συστήματα* (cancelable or revocable) – βιομετρικά συστήματα τα οποία αλλοιώνουν τα βιομετρικά χαρακτηριστικά με έναν μη αναστρέψιμο τρόπο τόσο κατά την καταχώρηση όσο και κατά την επερώτηση – έχουν εισαχθεί προκειμένουν να αντιμετωπίσουν την προστασία των βιομετρικών δεδομένων [92] [93] [94]. Πιο συγκεκριμένα, αντί να αποθηκεύεται η ψηφιακή αναπαράσταση του συγκεκριμένου βιολογικού χαρακτηριστικού, μια αλλοιωμένη εικόνα αυτου του χαρακτηριστικού αποθηκεύεται κατά τη φάση της καταχώρησης. Κάθε φορά που το ίδιο άτομο προσπαθεί να αποκτήσει πρόσβαση στο σύστημα, η συσκευή ανάγνωσης του βιομετρικού

χαρακτηριστικού αλλοιώνει την εικόνα και η διαδικασία της σύγκρισης περιλαμβάνει τη σύγκριση των δύο αλλοιωμένων εικόνων. Αυτή η τεχνική στόχο στηρίζει μια βασική όψη της ιδιωτικότητας, τη *μη συνδεσιμότητα*. Η τεχνική των ακυρώσιμων βιομετρικών χαρακτηριστικών προσθέτει όμως έναν πολύ σημαντικό περιορισμό. Έτσι, το τροποποιημένο βιομετρικό χαρακτηριστικό θα πρέπει όχι μόνο να μην ταιριάζει με το αρχικό βιομετρικό χαρακτηριστικό του ατόμου, αλλά και με κανένα άλλο βιομετρικό χαρακτηριστικό όλων των υπολοίπων ατόμων. Ο σημαντικός αυτός περιορισμός σε συνδυασμό με την έλλειψη υψηλών επιπέδων ακρίβειας των σημερινών τεχνικών [95] [96] καταδεικνύουν ότι απαιτείται ακόμη σημαντική έρευνα και προσπάθεια μέχρι τα ακυρώσιμα βιομετρικά συστήματα να καταφέρουν να πετύχουν τη μη συνδεσιμότητα.

Ο συνδυασμός πολλαπλών βιομετρικών χαρακτηριστικών (τα λεγόμενα πολυτροπικά βιομετρικά συστήματα) αποτελούν μια ακόμη προσπάθεια για προστασία της ιδιωτικότητας των ατόμων. Αν και το κύριο συγκριτικό τους πλεονέκτημα είναι ο πλούτος της πληροφορίας τον οποίο εκμεταλλεύονται, η υπεροχή τους δεν περιορίζεται εκεί. Έτσι, η ανάγκη για υποβολή διαφορετικών βιομετρικών χαρακτηριστικών στο βιομετρικό σύστημα για την πραγματοποίηση της ταυτοποίησης ή της επιβεβαίωσης ταυτότητας μειώνει την πιθανότητα για επίθεση στο πρώτο σημείο επίθεσης των βιομετρικών συστημάτων. Ο κύριος περιοριστικός παράγοντας για την εφαρμογή των πολυτροπικών βιομετρικών συστημάτων έγκειται στο κόστος τους, καθώς απαιτούν τον συνδυασμό ακριβού εξοπλισμού (για παράδειγμα, συσκευές ανάγνωσης ίριδας και δακτυλικών αποτυπωμάτων). Όμως, ένα σημαντικό βήμα προς την κατεύθυνση αυτή αποτελεί η εξαγωγή πολλαπλών βιομετρικών χαρακτηριστικών από το ίδιο σημείο του σώματος (για παράδειγμα από τα δάκτυλα των χεριών).

Μια ακόμη σημαντική εφαρμογή των βιομετρικών τεχνολογιών η οποία είναι στενά συνδεδεμένη με την ιδιωτικότητα είναι η *βιομετρική κρυπτογράφηση* [97]. Η κωδικοποίηση είναι η διαδικασία κατά την οποία η πληροφορία η οποία μεταδίδεται ή αποθηκεύεται είναι παραλλαγμένη. Η βασική ιδέα της κρυπτογραφίας βασισμένης σε κλειδί είναι ότι τα δεδομένα είναι κωδικοποιημένα έτσι ώστε μόνο ο αποστολέας και ο παραλήπτης τους να μπορούν να τα διαβάσουν. Συγκεκριμένα, στην κρυπτογραφία δημοσίου κλειδιού ο χρήστης έχει ένα ζεύγος κλειδιών, ένα ιδιωτικό και ένα δημόσιο κλειδί. Ο αποστολέας χρησιμοποιεί το δημόσιο κλειδί ως βάση για την κρυπτογράφηση των δεδομένων και μόνο ο κάτοχος του ιδιωτικού κλειδιού μπορεί να αποκρυπτογραφήσει τα κρυπτογραφημένα δεδομένα.

Η βιομετρική κρυπτογράφηση αποτελεί τη διαδικασία η οποία προκύπτει από τη συγχώνευση των βιομετρικών τεχνολογιών με την κρυπτογραφία. Σε μια προσπάθεια να εκμεταλλευτεί τα κύρια χαρακτηριστικά των βιομετρικών γνωρισμάτων, όπως είναι η μοναδικότητα και η ποικιλία, η βιομετρική κρυπρογράφηση χρησιμοποιεί ένα ή περισσότερα βιομετρικά γνωρίσματα ως μια μέθοδο για την ασφαλή διαχείριση κλειδιού. Με άλλα λόγια, η βιομετρική κρυπτογράφηση στοχεύει στο να βελτιώσει τα κρυπτογραφικά συστήματα έτσι ώστε τα κλειδιά να είναι λιγότερο ευάλωτα σε επιθέσεις μέσω της ασφαλούς σύνδεσης του κλειδιού με το βιομετρικό χαρακτηριστικό ώστε ούτε το κλειδί ουτε το βιομετρικό χαρακτηριστικό να μπορούν να ανακτηθούν από το αποθηκευμένο πρότυπο. Η διαρκώς αυαξόμενη ανταλλαγή πληροφορίας μέσω του Διαδικτύου έχει οδηγήσει σε μια εντεινόμενη ανάγκη για προστασία προσωπικών δεδομένων τα οποία είναι συνδεδεμένα σε ανοικτά δίκτυα και τα οποία είτε μεταδίδονται είτε αποθηκεύονται. Έτσι, η βιομετρική κρυπτογράφηση πηγαίνει ένα βήμα πιο πέρα από τα παραδοσιακά βιομετρικά συστήματα και επιτρέπει στα άτομα να χρησιμοποιούν ένα βιομετρικό γνώρισμα για πολλούς λογαριασμούς και σκοπούς δίχως τον φόβο ότι αυτοί οι ξεχωριστοί

ταυτοποιητές ή αυτές οι διαφορετικές χρήσεις θα συνδεθούν μεταξύ τους μέσω ενός βιομετρικού προτύπου αναφοράς [98].

Σύμφωνα με την Cavoukian [98], μια ακόμη εφαρμογή της βιομετρικής κρυπτογραφίας θα μπορούσε μια βάση δεδομένων η οποία είναι πολυτροπική περιέχοντας τόσο συμβατικά αλλά ανώνυμα πρότυπα αναφοράς όσο και ιδιωτικά πρότυπα αναφοράς τα οποία ελέγχουν έναν σύνδεσμο με τα κρυπτογραφημένα αρχεία του χρήστη. Έτσι η αποκωδικοποίηση των αρχείων του χρήστη θα είναι δυνατή αν επιτυγχάνεται θετικό και στα δύο είδη προτύπων αναφοράς. Η βιομετρική κρυπτογράφηση στοχεύει στο να προστατέψει την πληροφορία αυτήν κι έτσι μπορεί να δράσει ως υποστηρικτική τεχνολογία για ένα σύστημα προστασίας της ιδιωτικότητας.

6 *Προτεινόμενη Αρχιτεκτονική Βιομετρικών Συστημάτων και Καινοτόμος Μηχανισμός Συνδυασμού Αποτελεσμάτων*

Στο κεφάλαιο αυτό περιγράφεται η προτεινόμενη καινοτόμος αρχιτεκτονική βιομετρικών συστημάτων με στόχο τη βελτίωση της ακρίβειας των αποτελεσμάτων μέσω της εκμετάλλευσης μεγαλύτερου μέρους του πληροφοριακού πλούτου στο υπό ανάλυση ληφθέν βιομετρικό δείγμα και κατάλληλου έξυπνου συνδυασμού αυτού και παρουσιάζονται καινοτόμοι μηχανισμοί συνδυασμού των αποτελεσμάτων σύγκρισης των υπό ανάλυση βιομετρικών δεδομένων με τα βιομετρικά δεδομένα αναφοράς. Τέλος, παρατίθεται η αξιολόγηση αυτών των μηχανισμών.

6.1 *Προσδιορισμός του Προβλήματος και Προτεινόμενη Λύση*

Όπως είδαμε και στα προηγούμενα κεφάλαια, η απόδοση και η αξιοπιστία των υπαρχόντων βιομετρικών συστημάτων – ειδικά σε εφαρμογές μεγάλης κλίμακας – είναι ακόμη σε μη

ικανοποιητικά επίπεδο περιορίζοντας σημαντικά το εύρος των εφαρμογών τους. Για τον λόγο αυτό γίνονται πολλές ερευνητικές προσπάθειες σε διάφορα δομικά κομμάτια των βιομετρικών συστημάτων για περαιτέρω βελτίωσή τους. Έτσι, παρατηρείται διαρκής έρευνα για βελτιωμένες συσκευές ανάγνωσης βιομετρικών χαρακτηριστικών, για πιο εξελιγμένους αλγόριθμους εξαγωγής βιομετρικών χαρακτηριστικών, για πιο αποδοτική κατηγοριοποίηση και σύνθεση των τελευταίων (feature fusion) καθώς και για πιο αποτελεσματική σύνθεση των αποτελεσμάτων σύγκρισής τους (decision fusion).

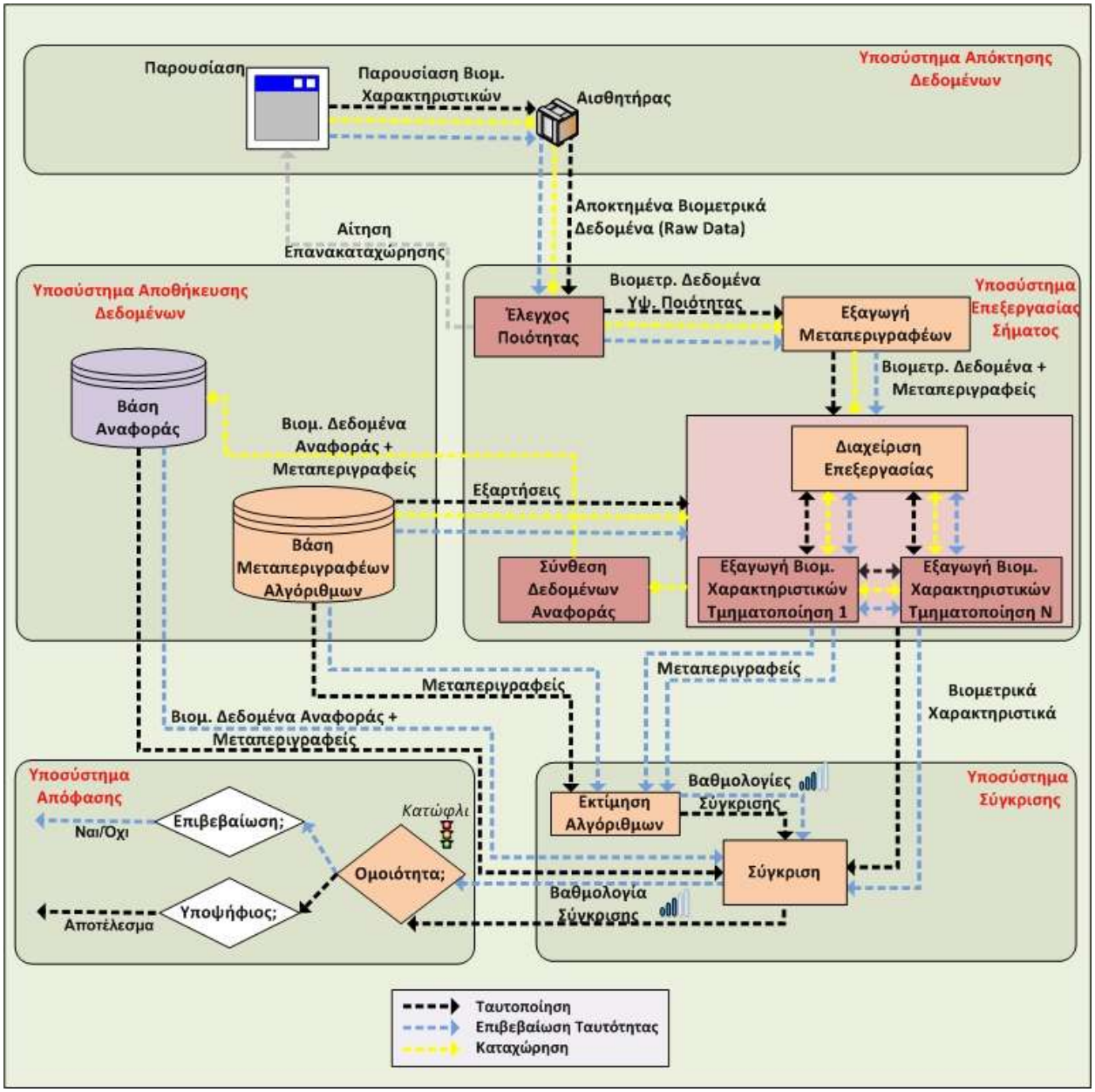
Όπως έχουμε ήδη δει, μια σημαντική προσπάθεια η οποία στοχεύει σε πιο αξιόπιστα βιομετρικά συστήματα αποτελούν και τα πολυτροπικά βιομετρικά συστήματα τα οποία υπόσχονται τη βελτίωση της απόδοσής τους για ταυτοποίηση και επιβεβαίωση ταυτότητας μέσω του συνδυασμού διαφορετικών βιομετρικών χαρακτηριστικών, όπως αναγνώριση προσώπου και φωνής. Τα βασικό μειονέκτημα των συστημάτων αυτών είναι κυρίως το κόστος τους, αλλά και η λειτουργία τους σε πραγματικό χρόνο.

Μέσα στο πλαίσιο αυτό, η ερευνητική μας προσπάθεια στράφηκε στην πρόταση μιας εξελιγμένης αρχιτεκτονικής μονοτροπικών βιομετρικών συστημάτων η οποία εκμεταλλεύται μεγαλύτερο μέρος του πλούτου των βιομετρικών δεδομένων από τις υπάρχουσες λύσεις χωρίς να επιβαρύνει με επιπρόσθετα κόστη όπως τα πολυτροπικά βιομετρικά συστήματα, ενώ η δυναμικότητα της προτεινόμενης λύσης επιτρέπει την προσαρμοστικότητα των συστημάτων και την υψηλότερη αξιοπιστία τους.

6.2 *Καινοτόμος Αρχιτεκτονική Βιομετρικών Συστημάτων*

Η βασική ιδέα της προτεινόμενης αρχιτεκτονικής έγκειται στο γεγονός ότι για κάθε βιομετρικό γνώρισμα έχει αναπτυχθεί μεγάλο πλήθος αλγορίθμων εξαγωγής και επεξεργασίας βιομετρικών

χαρακτηριστικών με καθέναν εκ των οποίων να αποδίδει καλύτερα υπό δεδομένες συνθήκες. Έτσι, για παράδειγμα στην περίπτωση των συστημάτων αναγνώρισης προσώπου κάποιοι αλγόριθμοι αποδίδουν καλύτερα υπό συνθήκες έντονης φωτεινότητας, άλλοι αλγόριθμοι είναι πιο αποτελεσματικοί για την αναγνώριση προσώπων συγκεκριμένης εθνικότητας, ενώ κάποιοι άλλοι εστιάζουν σε περιπτώσεις μερικού αποκλεισμού του προσώπου (partial occlusion). Στην προσπάθειά μας, λοιπόν, να συνθέσουμε ένα πιο γενικό βιομετρικό σύστημα (σε αντίθεση με τα περισσότερα υπάρχοντα βιομετρικά συστήματα τα οποία απαιτούν συγκεκριμένες συνθήκες λειτουργίας) οδηγηθήκαμε σε μια καινοτόμο αρχιτεκτονική η οποία παρουσιάζεται στην Εικόνα 12.



Εικόνα 12 Προτεινόμενη Εξελιγμένη Γενική Αρχιτεκτονική Βιομετρικού Συστήματος

Συγκρίνοντας την προτεινόμενη αρχιτεκτονική με την κυρίαρχη αρχιτεκτονική βιομετρικών συστημάτων η οποία και παρουσιάστηκε στην παράγραφο **Error! Reference source not found.**, παρατηρούμε ότι έχουμε εισαγάγει κάποια νέα δομικά στοιχεία στο υποσύστημα Επεξεργασίας Σήματος (*Εξαγωγή Μεταπεριγραφέων*, *Διαχείριση Επεξεργασίας*), στο υποσύστημα Αποθήκευσης

Δεδομένων (*Βάση Μεταπεριγραφέων Αλγόριθμων*) και στο υποσύστημα Σύγκρισης (*Εκτίμηση Αλγόριθμων*), ενώ

προκειμένου να είναι εφικτή η λειτουργικότητα της νέας αρχιτεκτονικής νέες διεπαφές σε υπάρχοντα δομικά στοιχεία έχουν προστεθεί, τα οποία και θα περιγραφούν συνοπτικά στη συνέχεια.

6.2.1 Εξαγωγή Μεταπεριγραφέων

Αυτό το δομικό στοιχείο στο υποσύστημα Επεξεργασίας Σήματος χρησιμοποιείται τόσο κατά τη φάση της καταχώρησης όσο και κατά τη φάση της επερώτησης. Η βασική του λειτουργία έγκειται στην επεξεργασία των βιομετρικών δεδομένων τα οποία αποστέλλονται από το υποσύστημα Απόκτησης Δεδομένων προκειμένουν να εξαχθούν κάποιοι *μεταπεριγραφείς* αυτών. Οι *μεταπεριγραφείς* αυτοί ουσιαστικά περιγράφουν τις γενικές συνθήκες των βιομετρικών δεδομένων. Έτσι, στην περίπτωση μιας εικόνας, περιγράφουν τα επίπεδα φωτεινότητας, την ύπαρξη μερικού αποκλεισμού, την πολυπλοκότητα του φόντου, ακόμη και το φύλο και την εθνικότητα με στόχο να προσφέρουν πολύτιμη πληροφορία σχετικά με την αποδοτικότητα των αλγόριθμων εξαγωγής χαρακτηριστικών. Η λειτουργία αυτή αποτελεί κεντρικό στοιχείο της προτεινόμενης αρχιτεκτονικής, καθώς αποτελεί τη βάση τόσο για το στάδιο σύνθεσης των βιομετρικών χαρακτηριστικών, όσο και για το στάδιο σύνθεσης των αποτελεσμάτων σύγκρισης και της τελικής απόφασης.

6.2.2 Διαχείριση Επεξεργασίας

Η *Διαχείριση Επεξεργασίας* στο υποσύστημα Επεξεργασίας Σήματος χρησιμοποιείται και στις δύο φάσεις των βιομετρικών συστημάτων και εστιάζει στη διαχείριση της επεξεργασίας των αλγορίθμων εξαγωγής χαρακτηριστικών. Λαμβάνοντας υπ’όψιν τις εξαρτήσεις των αλγόριθμων αναλαμβάνει τόσο να τροφοδοτήσει τους αλγόριθμους αυτούς με την είσοδο όσο και να

χρησιμοποιήσει κοινά τους σημεία για να βελτιώσει τα αποτελέσματά τους αλλά και την απόδοσή τους.

6.2.3 Βάση Μεταπεριγραφέων Αλγόριθμων

Στη βάση αυτή αποθηκεύονται οι μεταπεριγραφείς των αλγόριθμων εξαγωγής χαρακτηριστικών. Οι μεταπεριγραφείς αυτοί προσδιορίζουν μέσω συγκεκριμένων παραμέτρων αναλόγως με το είδος των βιομετρικών δεδομένων την αναμενόμενη αποτελεσματικότητα των αλγόριθμων αυτών. Έτσι, για παράδειγμα, στην περίπτωση συστήματος αναγνώρισης προσώπου, οι παράμετροι αυτοί αφορούν στην φωτεινότητα της εικόνας, στην ύπαρξη μερικών αποκλεισμών, σε συγκεκριμένα χαρακτηριστικά όπως γυαλιά και γένια, κλπ. Οι μεταπεριγραφείς αυτοί χρησιμοποιούνται από το υποσύστημα σύγκρισης προκειμένου να προσδιοριστεί ο βαθμός συμμετοχής του αποτελέσματος κάθε αλγόριθμου στο τελικό αποτέλεσμα αναλόγως με τις συνθήκες του υπό εξέταση βιομετρικού δείγματος.

6.2.4 Εκτίμηση Αλγόριθμων

Το δομικό στοιχείο αυτό αφορά στην εκτίμηση των αποτελεσμάτων των αλγόριθμων βάση των *μεταπεριγραφέων* τους οι οποίοι, όπως είδαμε, προσδιορίζουν το βαθμό απόδοσης των αλγόριθμων αυτών υπό ποικίλες συνθήκες και των *ειδικών συνθηκών του υπό εξέταση βιομετρικού δείγματος* όπως αυτές εξήχθησαν κατά το στάδιο Εξαγωγής Μεταπεριγραφέων. Έτσι, το δομικό στοιχείο αυτό «εκτιμά» την ακρίβεια των αποτελεσμάτων των αλγόριθμων και την οποία και παρέχει ως είσοδο στο δομικό στοιχείο «Σύγκριση» προκειμένου να συνδυαστούν πιο αποτελεσματικά τα αποτελέσματα των αλγόριθμων εξαγωγής χαρακτηριστικών και το προκύπτον αποτέλεσμα της σύγκρισης να είναι βασισμένο σε μεγαλύτερο μέρος του πληροφοριακού πλούτου του υπό εξέταση βιομετρικού δείγματος και συνεπώς πιο αξιόπιστο.

6.3 *Προτεινόμενος Μηχανισμός για αποτελεσματικό score-level fusion*

6.3.1 Το πρόβλημα

Τόσο τα πολυτροπικά όσο και τα διατροπικά βιομετρικά συστήματα (βλ. παράγραφο 3.3) αφορούν στον συνδυασμό (fusion) διαφορετικών πηγών προκειμένου να ξεπεραστούν οι περιορισμοί των μονοτροπικών βιομετρικών συστημάτων. Συνδυασμός μπορεί να λάβει χώρα σε τέσσερα επίπεδα πληροφορίας: αισθητήρα, χαρακτηριστικά, match score και επίπεδο απόφασης. Γενικώς προτιμάται το τρίτο επίπεδο πληροφορίας καθώς προσφέρει καλύτερη αντιστάθμιση μεταξύ του πληροφοριακού περιεχομένου και της συνδυαστικής ευκολίας. Ένας τέτοιος συνδυασμός αποτελεί πρόκληση για διάφορους λόγους, οι κυριότερες των οποίων είναι το γεγονός ότι οι βαθμολογίες διαφορετικών συγκριτών μπορεί να βασίζονται σε αποστάσεις ή μέτρα ομοιότητας, να ακολουθούν διαφορετικές κατανομές πιθανότητας, να παρέχουν διαφορετικά επίπεδα ακρίβειας καθώς και να είναι συσχετισμένες.

Υπάρχουν διάφορες τεχνικές για συνδυασμό βαθμολογιών οι οποίες γενικώς μπορούν να κατηγοριοποιηθούν σε τρεις κατηγορίες ως ακολούθως:

- Βασισμένη σε μετασχηματισμό: οι βαθμολογίες σύγκρισης πρώτα κανονικοποιούνται σε έναν κοινό χώρο και μετά συνδυάζονται. Η επιλογή του σχήματος κανονικοποίησης και τα βάρη εξαρτώνται από τα δεδομένα και απαιτούν σημαντική εμπειρική εκτίμηση.
- Βασισμένη σε κατηγοριοποιητή (classifier): Ο χειρισμός των βαθμολογιών από διαφορετικούς συγκριτές (matchers) γίνεται όπως ενός διανύσματος χαρακτηριστικών και ο κατηγοριοποιητής συντίθεται για τη διάκριση αυθεντικών και κίβδηλων βαθμολογιών (δηλαδή οι οποίες αντιστοιχούν σε χρήστη και σε μη χρήστη). Όταν ο συνδυασμός βιομετρικών βαθμολογιών αντιμετωπίζεται ως ένα πρόβλημα

κατηγοριοποίησης προκύπτουν ένα πλήθος προκλήσεων, όπως μη ισορροπημένο σύνολο εκπαίδευσης (το πλήθος των χρηστών είναι πολύ μικρότερο από το πλήθος των μη χρηστών), το κόστος της λανθασμένης κατηγοριοποίησης (αναλόγως με τον στόχο του συστήματος και το είδος της εφαρμογής είναι επιθυμητή η ελαχιστοποίηση του FAR ή του FRR) και η επιλογή του κατηγοριοποιητή.

- Βασισμένη σε πυκνότητα: αυτή η προσέγγιση βασίζεται στη δοκιμή λόγου πιθανοτήτων και απαιτεί προσέγγιση των πυκνοτήτων των χρηστών και των μη χρηστών. Το σημαντικό πλεονέκτημα αυτής της προσέγγισης είναι το γεγονός ότι επιτυγχάνει βέλτιστη απόδοση για ένα συγκεκριμένο επιλεγμένο σημείο, όπως είναι η τιμή του FAR, αν οι πυκνότητες των match scores είναι προσδιορισμένες με ακρίβεια.

Έστω ότι X = $[X_1, X_2, \dots, X_G]$ είναι τα match scores G διαφορετικών βιομετρικών συγκριτών, με X_g να είναι η τυχαία μεταβλητή η οποία αναπαριστά το match score του g-οστού συγκριτή, g=1,2,…,G και έστω ότι $f_{gen}(x)$ και $f_{imp}(x)$ είναι οι υπό συνθήκη joint συνδυασμένες πυκνότητας των G match scores δεδομένων των κλάσεων αυθεντικών και μη χρηστών αντιστοίχως, με x = $[x_1, x_2, \dots, x_G]$. Υποθέτουμε ότι καλούμαστε να αναθέσουμε το παρατηρηθέν διάνυσμα X των match scores στην κλάση των αυθεντικών ή μη χρηστών. Το πρόβλημά μας μπορεί να διατυπωθεί με τη μορφή στατιστικών δοκιμών. Έστω ότι Ψ είναι μια στατιστική δοκιμή της υπόθεσης H_0 «αντιστοιχεί σε έναν μη πιστοποιημένο χρήστη» έναντι της υπόθεσης H_1 «αντιστοιχεί σε έναν έγκυρο χρήστη» και Ψ(x) = i να δηλώνει ότι η απόφαση είναι υπέρ της H_i υπόθεσης, με i=0 ή 1. Όπως είδαμε και στην παράγραφο 4.3, η πιθανότητα απόρριψης της H_0 υπόθεσης, ενώ η τελευταία ισχύει είναι γνωστή ως FAR και αφορά στο μέγεθος ή στο επίπεδο της δοκιμής, ενώ η πιθανότητα απόρριψης της H_1 υπόθεσης ενώ η τελευταία ισχύει είναι γνωστή ως FRR. Η πιθανότητα ορθής απόρριψης της υπόθεσης H_0 όταν

ισχύει η υπόθεση H_1 δίνει *ορθώς αρνητικό* αποτέλεσμα και είναι γνωστό ως Genuine Accept Rate (GAR) και εκφράζει την ισχύ της δοκιμής.

Σύμφωνα με το θέωρημα Neyman-Pearson για δοκιμή της υπόθεσης H_0 έναντι της υπόθεσης H_1 υπάρχει μια δοκιμή Ψ και μια σταθερά η τέτοιες ώστε:

$$P(\Psi(\boldsymbol{X}) = 1|H_0) = \alpha \quad (33)$$

και

$$\Psi(x) = \begin{cases} 1, \ \frac{f_{gen}(x)}{f_{imp}(x)} \geq \eta \\ 0, \ \frac{f_{gen}(x)}{f_{imp}(x)} < \eta \end{cases} \quad (34)$$

Αν μια δοκιμή ικανοποιεί τις εξισώσεις (33) και (34) για κάποια τιμή του η, τότε αυτή αποτελεί την πιο ισχυρή δοκιμή για τη δοκιμή της υπόθεσης H_0 έναντι της υπόθεσης H_1 σε επίπεδο α.

Μεταφέροντας το θεώρημα αυτό στο πρόβλημά μας, δεδομένης μιας τιμής του FAR α, η βέλτιστη δοκιμή για την απόφαση του αν το διάνυσμα X του score αντιστοιχεί σε έναν γνήσιο χρήστη ή σε έναν μη έγκυρο χρήστη είναι η δοκιμή λόγου πιθανοτήτων, όπως αυτή δίνεται από την εξίσωση (34). Έτσι για μια δεδομένη τιμή FAR στην οποία στοχεύουμε με το βιομετρικό μας σύστημα, μπορούμε να επιλέξουμε ένα τέτοιο κατώφλι η τέτοιο ώστε η δοκιμή λόγου πιθανοτήτων να μεγιστοποιεί την τιμή του GAR. Βασιζόμενοι στο ανωτέρω θεώρημα, είμαστε βέβαιοι ότι δεν υπάρχει άλλος κανόνας απόφασης ο οποίος να δίνει μεγαλύτερη τιμή του GAR. Βασική προϋπόθεση όμως να ισχύει το παραπάνω είναι να είναι γνωστές οι πυκνότητες των $f_{gen}(x)$ και $f_{imp}(x)$. Στην πράξη αυτό μπορεί να λάβει χώρα μέσω της εκπαίδευσης του μηχανισμού με ένα σύνολο δεδομένων match scores τόσο γνήσιων όσο και μη χρηστών του

συστήματος, με την τελική αποτελεσματικότητα του μηχανισμού να εξαρτάται σημαντικά από την ακρίβεια των εκτιμήσεων αυτών.

Οι Nandakumar et al [103] παρουσίασαν μια εργασία τους η οποία εστίαζε στην μοντελοποίηση των πυκνοτήτων αυτών των match scores μέσω Gaussian Mixture Models (GMM) και στην εφαρμογή αυτών σε τρεις πολυτροπικές βάσεις βιομετρικών δεδομένων οι οποίες συμπεριλάμβαναν δεδομένα για δακτυλικά αποτυπώματα, ίριδα και πρόσωπο. Ο λόγος που βασίστηκαν στα GMM είναι ότι έχουν εφαρμοστεί επιτυχώς για τον προσδιορισμό αυθαίρετων πυκνοτήτων με τα θεωρητικά αποτελέσματα να δείχνουν ότι οι εκτιμήσεις των πυκνοτήτων με χρήση πεπερασμένων μοντέλων μίξης (finite mixture models) πράγματι συγκλίνουν στην πραγματική πυκνότητα όταν είναι διαθέσιμος ένας ικανός αριθμός δειγμάτων εκπαίδευσης (training samples). Η επιλογή του κατάλληλου αριθμού παραγόντων για τα GMM μοντέλα βασίστηκε στον αλγόριθμο EM (Expectation Maximization) [104]. Τα αποτελέσματα της έρευνάς τους για την ελαχιστοποίηση του FRR για δεδομένο FAR συγκρίθηκαν με τον Support Vector Machine (SVM) κατηγοριοποιητή, καθώς και με μια συνήθη τεχνική βασισμένη σε μετασχηματισμούς, το άθροισμα των βαθμολογιών επιτυγχάνοντας σε γενικές γραμμές υψηλότερες τιμές επιτυχούς ταυτοποίησης.

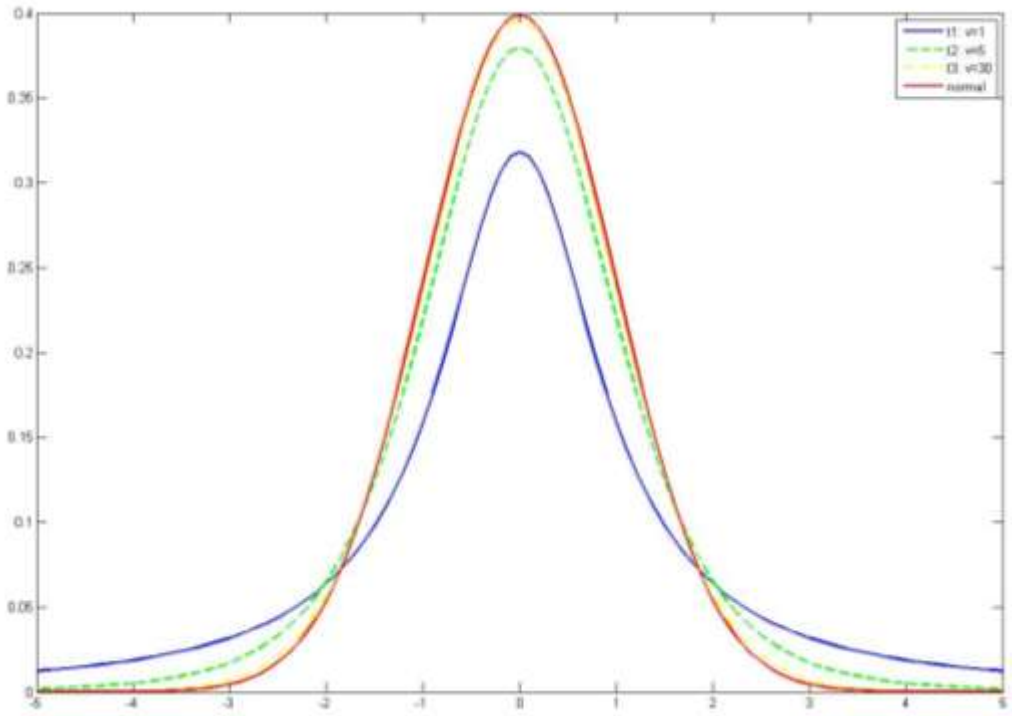
6.3.2 Προτεινόμενος Καινοτόμος Μηχανισμός

Η έρευνα μας βασίστηκε στο γεγονός ότι παρ'όλο που τα GMM είναι γενικώς αρκετά αποδοτικά όσον αφορά στην προσέγγιση της πυκνότητας των βαθμολογιών χρηστών και μη, συνοδεύονται από κάποια προβλήματα. Το κυριότερο αυτών αφορά στις «ουρές» των κατανομών αυτών οι οποίες πολύ συχνά είναι πιο μικρές απ'ό,τι χρειάζεται. Έτσι, σε πολυδιάστατα προβλήματα, όπως είναι η πιστοποίηση της ταυτότητας χρηστών σε βιομετρικά συστήματα, το θέμα της προστασίας έναντι των outliers είναι αρκετά σημαντικό και γενικώς

δυσεπίλυτο. Για τον λόγο αυτόν η έρευνά μας εστιάζει στην εφαρμογή μοντέλων μίξης Student t κατανομών.

6.3.2.1 Η κατανομή Student t και η Ομοιόμορφη κατανομή

Η Student t (ή απλώς t) κατανομή παρέχει μια εναλλακτική της κανονικής κατανομής με μεγαλύτερη «ουρά». Έτσι, αποτελεί μια πιο εύρωστη προσέγγιση καθώς οι παρατηρήσεις οι οποίες είναι μη κοινές για έναν παράγοντα λαμβάνουν μειωμένο βάρος κατά τον υπολογισμό των παραμέτρων. Επίσης, η χρήση των t παραγόντων δίνει λιγότερο υπερβολικές εκτιμήσεις των posterior πιθανοτήτων της συμμετοχής των παραγόντων του μοντέλου μίξης [105]. Στην προσέγγιση αυτή, η κανονική κατανομή κάθε παράγοντα στο μοντέλο ενσωματώνεται σε μια ευρύτερη κλάση ελλειπτικών συμμετρικών κατανομών με μια επιπρόσθετη παράμετρο η οποία καλείται βαθμοί ελευθερίας *ν*. Όταν το ν τείνει στο άπειρο, τότε η t κατανομή προσεγγίζει την κανονική κατανομή, όπως μπορούμε να παρατηρήσουμε και στην Εικόνα 13.



Εικόνα 13 Γραφική παράσταση Student t κατανομών με δείκτες ελευθερίας v=1, 5 και 30 σε αντιστοιχία με την κανονική κατανομή. Παρατηρούμε ότι καθώς το v αυξάνεται, η Student t κατανομή τείνει στην κανονική κατανομή.

Μια σημαντική εφαρμογή των κανονικών μοντέλων μίξης απαντάται στον χώρο της ανάλυσης clusters [106]. Πέραν της μαθηματικής της βάσης, η προσέγγιση αυτή δεν περιορίζεται στην παραγωγή σφαιρικών clusters βασιζόμενα στην Ευκλείδια απόσταση όπως οι αλγόριθμοι k-μέσου σε αντίθεση με την απόσταση Mahalanobis η οποία επιτρέπει για συσχετίσεις εντός του cluster μεταξύ μεταβλητών στο διάνυσμα χαρακτηριστικών *Y*, ενώ σε σχέση με άλλες τεχνικές στον ίδιο χώρο οι οποίες βασίζονται αποκλειστικά στην απόσταση Mahalanobis, αυτού του είδους το clustering λαμβάνει υπ'όψιν τον παράγοντα κανονικοποίησης $|\Sigma_i|^{-1/2}$ στην εκτίμηση της πολυμεταβλητής κανονικής πυκνότητας η οποια υιοθετείται για την κατανομή του παράγοντα του *Y* ο οποίος αντιστοιχεί στο i-οστό cluster. Αν και ακόμη και μια πρόχειρη εκτίμηση του πίνακα συνδιακύμανσης (covariance matrix) Σ_i συχνά επαρκεί για το clustering, η επιρροή των outliers μπορεί να είναι πολύ σημαντική.

6.3.2.2 Πολυμεταβλητή κατανομή Student t

Έστω ότι $\boldsymbol{X} = [x_1, x_2, \ldots, x_n]$ είναι ένα παρατηρηθέν p-διάστατο τυχαίο δείγμα μεγέθους *n*. Αν ακολουθήσουμε την προσέγγιση μοντέλου μίξης κανονικών κατανομών προκειμένου να εξαγάγουμε συμπεράσματα από τα δεδομένα αυτά, κάθε σημείο θεωρείται ότι είναι στιγμιότυπο του τυχαίου *p*-διάστατου διανύσματος $\boldsymbol{X}$ με τη *M*-παραγόντων συνάρτηση πυκνότητας πιθανότητας Μ-παραγόντων μίξης κανονικών κατανομών να εκφράζεται με την ακόλουθη σχέση:

$$f(\boldsymbol{x};\boldsymbol{\Psi}) = \sum_{i=1}^{M} \pi_i \varphi(\boldsymbol{x};\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i) \quad (1)$$

με π_i να είναι τα βάρη που καθορίζουν το ποσοστό συμμετοχής της κάθε κατανομής στη μίξη για τα οποία ισχύει :

$$\sum_{i=1}^{M} \pi_i = 1 \ (2)$$

με την

$$\varphi(\boldsymbol{x};\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i) = \frac{1}{(2\pi)^{-p/2}|\Sigma_i|^{-1/2}} \times \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^T \Sigma_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)\right\} \quad (3)$$

να εκφράζει την p-διάστατη πολυμεταβλητή κανονική συνάρτηση πυκνότητας πιθανότητας με μέση τιμή μ_i και πίνακα συνδιακύμανσης Σ_i.

και με $\boldsymbol{\Psi}$ να αποτελεί το σύνολο των παραμέτρων του μοντέλου, δηλαδή $\boldsymbol{\Psi}=(\pi_1, \pi_2, \ldots, \pi_M, \boldsymbol{\theta}^T)$, με το $\boldsymbol{\theta}$ να συνιστάται από τις μέσες τιμές μ_i και τις συνδιακυμάνσεις Σ_i, *i = 1,..., M.*

Ένας τρόπος για να γίνει πιο ευρεία αυτή η παραμετρική οικογένεια για πιθανά outliers ή για δεδομένα με μεγαλύτερες ουρές από τις συνήθεις είναι η υιοθέτηση ενός mixture model δύο παραγόντων :

$$(1\text{-}\varepsilon)\ \varphi(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) + \ \varepsilon\varphi(\boldsymbol{x};\boldsymbol{\mu},c\boldsymbol{\Sigma}) \ (4)$$

με c να είναι ένας μεγάλος αριθμός και ε ένας μικρός αριθμός, αναπαριστώντας τη μικρή αναλογία παρατηρήσεων οι οποίες έχουν σχετικά μεγάλη διακύμανση.

Το παραπάνω κανονικό scale mixture μοντέλο μπορεί να γραφεί ως εξής:

$$\int \varphi(\boldsymbol{x}; \boldsymbol{\mu}; \boldsymbol{\Sigma}/u)\, dH(u) \quad (5)$$

με Η να είναι η κατανομή πιθανότητας η οποία τοποθετεί τον παράγοντα με συμμετοχή (1-ε) στο σημείο u=1 και για τον παράγοντα με συμμετοχή ε στο σημείο u=1/c. Αν αντικαταστήσουμε το Η με τη συνάρτηση πυκνότητας πιθανότητας μιας x^2 τυχαίας μεταβλητής με *ν* βαθμούς ελευθερίας, δηλαδή με την τυχαία μεταβλητή U με κατανομή γάμμα:

$$\text{U} \sim \text{gamma}\left(\frac{v}{2}, \frac{v}{2}\right)$$

Η συνάρτηση πυκνότητας πιθανότητας της Γάμμα κατανομής εκφράζεται ως:

$$f(u; a, b) = \frac{\beta^{\alpha} u^{\alpha-1}}{\Gamma(\alpha)} \times \exp(-\beta u)\, I_{(0,\infty)}(u) \text{ με } (\alpha,\beta)>0 \quad (6)$$

και τη συνάρτηση $I_{(0,\infty)}(u)$ να ορίζεται ως ακολούθως:

$$I_{(0,\infty)}(u) = \begin{cases} 1, & u > 0 \\ 0, & u \leq 0 \end{cases} \quad (7)$$

Η t κατανομή λαμβάνεται τότε με location παράμετρο *μ*, θετικό πίνακα εσωτερικού εσωτερικού γινομένου *Σ* και *ν* βαθμούς ελευθερίας:

$$f(\boldsymbol{x};\ \mu, \boldsymbol{\Sigma}, v) = \frac{\Gamma\left(\frac{v+p}{2}\right)\ |\boldsymbol{\Sigma}|^{-1/2}}{(\pi v)^{\frac{1}{2}p}\Gamma\left(\frac{v}{2}\right)\left\{1 + \frac{\delta(\boldsymbol{x}, \mu; \boldsymbol{\Sigma})}{v}\right\}^{\frac{1}{2}(v+p)}} \quad (8)$$

με τη συνάρτηση δ να ορίζεται ως ακολούθως:

$$\delta(\boldsymbol{x}, \mu; \boldsymbol{\Sigma}) = (\boldsymbol{x} - \mu)^{\boldsymbol{T}} \boldsymbol{\Sigma}^{-\mathbf{1}} (\boldsymbol{x} - \mu) \quad (9)$$

και να αφορά στην απόσταση Mahalanobis (στο τετράγωνο) μεταξύ των $\boldsymbol{x}$ και $\boldsymbol{\mu}$. Αν το $\nu > 1$, το μ είναι η μέση τιμή του $\boldsymbol{X}$, ενώ αν $\nu>2$ ο πίνακας συνδιακύμανσης είναι ο $\nu(\nu-2)^{-1}\Sigma$. Για ν τείνον στο άπειρο το U τείνει στη μονάδα με πιθανότητα ίση με τη μονάδα κι έτσι το $\boldsymbol{X}$ γίνεται οριακά πολυμεταβλητή κανονική με μέσο μ και πίνακα συνδιακύμανσης $\boldsymbol{\Sigma}$. Η οικογένεια των t κατανομών έτσι παρέχει μια εναλλακτική με μεγαλύτερες «ουρές» από την κανονική κατανομή με μέση τιμή μ και πίνακα συνδιακυμάνσεων ίσο με ένα μονοδιάστατο πολλαπλάσιο του $\boldsymbol{\Sigma}$ για $\nu>2$.

6.3.2.3 ΕΜ Αλγόριθμος για μοντέλο μίξης Student t

Προκειμένου να βρεθεί το πλήθος των παραγόντων τα οποία θα χρησιμοποιηθούν αναλόγως με τα δεδομένα εκπαίδευσης αλλά και τις ανάγκες του εκάστοτε προβλήματος καθώς και οι παράμετροι των επιμέρους κατανομών, προσαρμόστηκε ο ΕΜ αλγόριθμος για g-παραγόντων μείγμα πολυμεταβλητών t κατανομών [106]. Η εργασία αυτή των Peel και McLachlan είναι βασιμένη στην εφαρμογή του αλγόριθμου αυτού για ML (Maximum Likelihood) προσέγγιση στην περίπτωση ενός παράγοντα t κατανομής από τους McLachlan και Krishnan [107].

Έτσι, για την εφαρμογή του ΕΜ αλγόριθμου, το πλήρες διάνυσμα δεδομένων είναι :

$$x_c = (x_0^T, z_1^T, \dots, z_n^T, u_1, \dots, u_n)^T \quad (10)$$

με $x_0 = (x_1^T, \dots, x_n^T)^T$ να αφορά στο διάνυσμα των παρατηρηθέντων δεδομένων, $z_1^T, \dots, z_n^T$ να είναι τα component-label διανύσματα τα οποία ορίζουν τον παράγοντα προέλευσης των $x_1^T, \dots, x_n^T$ αντιστοίχως και

$$z_{ij} = (z_j)_i = \begin{cases} 1, & x_j \in i\ component \\ 0, & x_j \notin i\ component \end{cases} \quad (11)$$

Στα πλαίσια της *t* κατανομής, θεωρούμε ότι τα παρατηρηθέντα δεδομένα επηυξημένα με τα z_j ως μη πλήρη και εισάγονται στο διάνυσμα πλήρων δεδομένων τα επιπρόσθετα δεδομένα, $u_1^T, \dots, u_n^T$ τα οποία και ορίζονται έτσι ώστε δεδομένου ότι $z_{ij} = 1$,

$$\boldsymbol{X}_j \,|u_j, z_{ij} = 1 \sim N\left(\mu_i, \frac{\Sigma_i}{u_j}\right) \text{(12)}$$

ανεξάρτητα για j=1,…,n και

$$\mathrm{U_j}|\ \mathrm{z_{ij}} = 1 \sim \mathrm{gamma}\left(\frac{1}{2}\mathrm{v_i}, \frac{1}{2}\mathrm{v_i}\right) \text{(13)}$$

Σύμφωνα με την τελευταία σχέση, με δεδομένα τα $z_1, \dots, z_n$, οι $U_1, \dots, U_n$ είναι ανεξάρτητες. Η πιθανότητα των πλήρων δεδομένων $L_c(\boldsymbol{\Psi})$ μπορεί να παραγοντοποιηθεί στο γινόμενο των οριακών πυκνοτήτων των $\boldsymbol{Z}_j$, των υπό συνθήκη πυκνοτήτων των $\boldsymbol{U}_j$ δεδομένων των z_j και των υπό συνθήκη πυκνοτήτων των $\boldsymbol{X}_j$ δεδομένων των u_j και z_j. Αντιστοίχως η λογαριθμική πιθανότητα των πλήρων δεδομένων μπορεί να εκφραστεί ως:

$$logL_c(\boldsymbol{\Psi}) = logL_{1c}(\pi) + logL_{2c}(v) + logL_{3c}(\theta) \text{ (14)}$$

με

$$logL_{1c}(\pi) = \sum_{i=1}^{g}\sum_{j=1}^{n} z_{ij}\ log\pi_i \text{ (15), με } \pi = (\pi_1, \dots, \pi_g)^T$$

και

$$logL_{2c}(v) = \sum_{i=1}^{g}\sum_{j=1}^{n} z_{ij}\ \left\{-log\Gamma\left(\frac{1}{2}v_i\right) + \frac{1}{2}v_i log\left(\frac{1}{2}v_i\right) + \frac{1}{2}v_i(log\ u_i - u_i) - \log u_j\right\} \text{(16)}$$

και

$$logL_{3c}(\theta) = \sum_{i=1}^{g}\sum_{j=1}^{n} z_{ij}\ \left\{-\frac{1}{2}p\ log\Gamma(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}_i| - \frac{1}{2}u_i\left(x_j - \mu_i\right)^T \boldsymbol{\Sigma}_i^{-1}\left(x_j - \mu_i\right)\right\} \text{(17)}$$

6.3.2.3.1 Το βήμα E (Expectation)

Στην (k+1)-οστή επανάληψη του ΕΜ αλγόριθμου στο Ε-βήμα απαιτείται ο υπολογισμός του $Q(\boldsymbol{\Psi},\boldsymbol{\Psi}^{(k)})$, της τρέχουσας υπό συνθήκη προσδοκίας για τη συνάρτηση λογαριθμικής πιθανότητας των πλήρων δεδομένων $logL_c(\boldsymbol{\Psi})$. Το Ε-βήμα μπορεί να επηρεαστεί λαμβάνοντας πρώτα την προσδοκία του $logL_c(\boldsymbol{\Psi})$ υπό συνθήκη επίσης επί των $z_1, \dots, z_n$ και επί του x_0 και τέλος επί των z_j δεδομένου του x_0. Από τις τελευταίες σχέσεις ((15)-(17)) προκύπτει οτί τα παραπάνω απαιτούν τους υπολογισμούς των:

$$E_{\boldsymbol{\Psi}^{(k)}}(Z_{ij}\,|x_j),\ E_{\boldsymbol{\Psi}^{(k)}}(U_j\,|x_j, z_j),\ E_{\boldsymbol{\Psi}^{(k)}}(logU_j\,|x_j, z_j)\ \text{για i=1,…,g και j=1,…,n (18)}$$

Προκύπτει ότι

$$E_{\boldsymbol{\Psi}^{(k)}}(Z_{ij}\,|x_j) = \tau_{ij}^{(k)},\ (19)$$

με

$$\tau_{ij}^{(k)} = \frac{\pi_i^{(k)} f\left(x_j; \mu_i^{(k+1)}, \boldsymbol{\Sigma}_i^{(k+1)}, v_i^{(k+1)}\right)}{f(x_j; \boldsymbol{\Psi}^{(k+1)})}\ (20)$$

να είναι η posterior πιθανότητα ότι το x_j ανήκει στον i-οστό παράγοντα της μίξης, κάνοντας χρήση του τρέχοντος ταιριάσματος $\boldsymbol{\Psi}^{(k)}$ για $\boldsymbol{\Psi}$ (i=1,…,g, j=1,…,n). Δεδομένου ότι η γάμμα κατανομή είναι η συζυγής prior κατανομή του U_j, εύκολα προκύπτει ότι η υπό συνθήκη κατανομή του τελευταίου δεδομένων των $\boldsymbol{X}_j = x_j$ και $Z_{ij} = 1$ είναι

$$U|x_j, z_{ij} = 1 \sim gamma(m_{1i}, m_{2i}),\ (21)$$

με $m_{1i} = \frac{1}{2}(v_i + p)$ (22) και $m_{1i} = \frac{1}{2}\{v_i + \delta(x_j, \mu_i; \boldsymbol{\Sigma}_i)\}$ (23).

Από την (21) προκύπτει ότι :

$$E\left(U_j \middle| x_j, z_{ij} = 1\right) = \frac{v_i + p}{v_i + \delta(x_j + \mu_i; \boldsymbol{\Sigma}_i)}$$

οπότε

$$E_{\boldsymbol{\Psi}^{(k)}}\left(U_j \middle| x_j, z_{ij} = 1\right) = \frac{v_i^{(k)} + p}{v_i^{(k)} + \delta(x_j, \mu_i^{(k)}; \boldsymbol{\Sigma}_i^{(k)})} = u_{ij}^{(k)}$$

Ο υπολογισμός της σχέσης (19) θα βασιστεί στο γεγονός ότι αν η τυχαία μεταβλήτη W έχει μια γάμμα κατανομή (gamma(α,β)) τότε $E(\log W) = \psi(\alpha) - \log \beta$, με $\psi(\alpha) = \left(\left\{\frac{\vartheta \Gamma(s)}{\vartheta s}\right\} / \Gamma(s)\right)_{s=a}$ να είναι η Digamma κατανομή. Καταλήγει λοιπόν στη σχέση:

$$E_{\boldsymbol{\Psi}^{(k)}}\left(U_j \middle| x_j, z_{ij} = 1\right) = \psi\left(\frac{v_i^{(k)}+p}{2}\right) - log\left\{\frac{1}{2}\left(v_i^{(k)} + \delta(x_j, \mu_i^{(k)}; \boldsymbol{\Sigma}_i^{(k)})\right)\right\} \Longrightarrow$$

$$E_{\boldsymbol{\Psi}^{(k)}}\left(U_j \middle| x_j, z_{ij} = 1\right) = \log u_{ij}^{(k)} + \left\{\psi\left(\frac{v_i^{(k)}+p}{2}\right) - log\left(\frac{v_i^{(k)}+p}{2}\right)\right\} \quad (24)$$

Από τις παραπάνω σχέσεις προκύπτει από τη (14):

$$Q\left(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(k)}\right) = Q_1\left(\pi; \boldsymbol{\Psi}^{(k)}\right) + Q_2\left(v; \boldsymbol{\Psi}^{(k)}\right) + Q_3\left(\theta; \boldsymbol{\Psi}^{(k)}\right), \quad (25)$$

με

$$Q_1\left(\pi; \boldsymbol{\Psi}^{(k)}\right) = \sum_{i=1}^{g}\sum_{j=1}^{n} \hat{\tau}_{ij}^{(k)} \log \pi_i \quad (26)$$

$$Q_2\left(v; \boldsymbol{\Psi}^{(k)}\right) = \sum_{i=1}^{g}\sum_{j=1}^{n} \hat{\tau}_{ij}^{(k)} Q_{2j}\left(v_i; \boldsymbol{\Psi}^{(k)}\right) \quad (27)$$

$$Q_3\left(\theta; \boldsymbol{\Psi}^{(k)}\right) = \sum_{i=1}^{g}\sum_{j=1}^{n} \hat{\tau}_{ij}^{(k)} Q_{3j}\left(\theta_i; \boldsymbol{\Psi}^{(k)}\right) \quad (28)$$

Αγνοώντας τους όρους οι οποίοι δεν εμπεριέχουν v_i λαμβάνουμε:

$$Q_{2j}\left(v_i; \boldsymbol{\Psi}^{(k)}\right) = -\log\Gamma\left(\frac{1}{2}v_i\right) + \frac{1}{2}v_i \log\left(\frac{1}{2}v_i\right) + \frac{1}{2}v_i \left\{\sum_{j=1}^{n}\left(log\, u_{ij}^{(k)} - u_{ij}^{(k)}\right) + \right.$$

$$\left. + \psi\left(\frac{v_i^{(k)}+p}{2}\right) - log\left(\frac{v_i^{(k)}+p}{2}\right)\right\} \quad (27)$$

και

$$Q_{3j}\left(\theta_i; \boldsymbol{\Psi}^{(k)}\right) = -\frac{1}{2}\, p \log\,(2\pi) - \frac{1}{2}\log|\Sigma_i| + \frac{1}{2}\, p \log\; u_{ij}^{(k)} - \frac{1}{2}\; u_{ij}^{(k)}\left(x_j - \mu_i\right)^T \boldsymbol{\Sigma}_i^{-1}\left(x_j - \mu_i\right)$$

(28)

6.3.2.3.2 Το βήμα M (Maximisation)

Στο M-βήμα της (k+1) επανάληψης του ΕΜ αλγόριθμου προκύπτει από την (25) ότι οι όροι $\pi^{(k+1)}$, $\theta^{(k+1)}$ και $v^{(k+1)}$ μπορούν να υπολογιστούν ανεξάρτητα μέσω των (26), (27) και (28) με τις λύσεις των $\pi^{(k+1)}$ και $\theta^{(k+1)}$ να είναι σε κλειστή μορφή και μόνο τις νέες τιμές για το $v^{(k+1)}$ να υπολογίζονται μέσω επαναλήψεων.

Έτσι, οι συντελεστές αναλογίας στη μίξη υπολογίζονται λαμβάνοντας υπ'όψιν τον πρώτο όρο $Q_1\left(\pi; \boldsymbol{\Psi}^{(k)}\right)$, με βάση το οποίο ο όρος $\pi^{(k+1)}$ να δίνεται από το μέσο όρο των posterior πιθανοτήτων της συμμετοχής των παραγόντων στη μίξη,

$$\pi^{(k+1)} = \sum_{j=1}^{n} \tau_{ij}^{(k)}/n \;\; (i = 1, \ldots, g) \quad (29)$$

Προκειμένου να υπολογιστούν οι νέες τιμές των μ_i και Σ_i πρέπει να λάβουμε υπ'όψιν μας τον όρο

$$Q_3\left(\theta_i; \boldsymbol{\Psi}^{(k)}\right)$$

ο οποίος ουσιαστικά αντιστοιχεί στη λογαριθμική συνάρτηση πιθανότητας n ανεξάρτητων παρατηρήσεων $x_1, \ldots, x_n$ με κοινή μέση τιμή μ_i και πίνακες συνδιακύμανσης Σ_i/u_1^k, …, Σ_i/u_n^k αντιστοίχως. Είναι έτσι ισοδύναμο με τον υπολογισμό της μέσης τιμής δείγματος με βάρη και του πίνακα δυνδιακύμανσης του δείγματος με βάρη $u_1^k, \ldots, u_n^k$, δηλαδή:

$$\mu_i^{(k+1)} = \sum_{j=1}^{n} \tau_{ij}^{(k)}\, u_{ij}^{(k)} x_j \,/\, \sum_{j=1}^{n} \tau_{ij}^{(k)}\, u_{ij}^{(k)} \quad (30)$$

και

$$\boldsymbol{\Sigma}_i^{(k+1)} = \sum_{j=1}^{n} \tau_{ij}^{(k)}\, u_{ij}^{(k)} (x_j - \mu_i^{(k+1)})(x_j - \mu_i^{(k+1)})^T \,/\, \sum_{j=1}^{n} \tau_{ij}^{(k)} \quad (31)$$

Όπως μπορεί να αποδειχθεί επιλέγει $\mu_i^{(k+1)}$ και $\Sigma_i^{(k+1)}$ αποτελεσματικά μέσω προσέγγισης ελαχίστων τετραγώνων με βάρη. Το Ε-βήμα ενημερώνει τα βάρη $u_{ij}^{(k)}$, ενώ το Μ-βήμα επιλέγει αποτελεσματικά τα $\mu_i^{(k+1)}$ και $\boldsymbol{\Sigma}_i^{(k+1)}$ όπως είπαμε. Από τη μορφή της (30) προκύπτει ότι όσο το $v_i^{(k)}$μειώνεται, τόσο ο βαθμός μείωσης του βάρους ενός outlier μεγαλώνει. Για πεπερασμένο $v_i^{(k)}$, τείνοντος του $||x_j||$ προς το άπειρο, η επίδραση στην εκτίμηση της location παραμέτρου του i-οστού παράγοντα τείνει στο μηδέν, ενώ η επίδραση στην εκτίμηση του scale του ίδιου παράγοντα παραμένει φραγμένη αλλά δεν εξαφανίζεται.

Ακολουθώντας την προτεινόμενη προσέγγιση των Kent, Tyler και Vardi [108] στην περίπτωση *t* κατανομής ενός παράγοντα, οι Peel και McLachlan [106] προτείνουν την αντικατάσταση του διαιρέτη $\sum_{j=1}^{n} \tau_{ij}^{(k)}$ στη σχέση (31) με τον όρο:

$$\sum_{j=1}^{n} \tau_{ij}^{(k)} u_{ij}^{(k)}$$

Αποδεικνύεται ότι αν οι βαθμοί ελευθερίας v_i είναι δεδομένοι εξ αρχής για κάθε παράγοντα, τότε το Μ-βήμα είναι σε κλειστή μορφή. Στην περίπτωση αυτή, η εκτίμηση των παραμέτρων των μοντέλων είναι μια μορφή της Μ-εκτίμησης. Όμως, μια ελκυστικό στοιχείο της χρήσης της *t* κατανομής για τη μοντελοποίηση των κατανομών των παραγόντων είναι ότι οι βαθμοί ευρωστίας όπως αυτοί ελέγχονται από τους όρους v_i μπορούν να εξαχθούν από τα δεδομένα μέσω του υπολογισμού της ML (Maximisation Likelihood) εκτίμησής τους. Στην περίπτωση αυτή, γίνεται υπολογισμός επίσης στο Μ-βήμα της νέας εκτίμησης του $v_i^{(k+1)}$ του v_i. Υπολογίζοντας το αριστερό μέρος της εξίσωσης

$$\sum_{\boldsymbol{j=1}}^{\boldsymbol{n}} \vartheta Q_{2j}(v_i; \boldsymbol{\Psi}^{(k)})/\vartheta v_i = 0$$

προκύπτει ότι το $v_i^{(k+1)}$ αποτελεί τη λύση της εξίσωσης :

$$\left\{-\psi\left(\frac{1}{2}v_i\right) + log\left(\frac{1}{2}v_i\right) + 1 + \frac{1}{n_i^{(k)}}\sum_{j=1}^{n}\tau_{ij}^{(k)}\left(\log u_{ij}^{(k)} - u_j^{(k)}\right) + \psi\left(\frac{\left(v_i^{(k)}+p\right)}{2}\right) -\right.$$

$$logvik+p2=0 \quad (32)$$

με $n_i^{(k)} = \sum_{j=1}^{n}\tau_{ij}^{(k)}$.

6.3.2.4 Εφαρμογή μοντέλων μίξης Student t κατανομών

Έστω ότι $\varphi^G(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{v})$ είναι η G-μεταβλητή Student t πυκνότητα με διάνυσμα μέσων τιμών $\boldsymbol{\mu}$, πίνακα συνδιακύμανσης $\boldsymbol{\Sigma}$ και διάνυσμα βαθμών ελευθερίας $\boldsymbol{v}$. Τότε έχουμε ότι :

$$\varphi^G(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{v}) = \frac{\Gamma\left(\frac{v+G}{2}\right)\ |\boldsymbol{\Sigma}|^{-1/2}}{(\pi v)^{\frac{1}{2}G}\Gamma\left(\frac{v}{2}\right)\left\{1+\frac{\delta(\boldsymbol{x},\mu;\boldsymbol{\Sigma})}{v}\right\}^{\frac{1}{2}(v+G)}}$$

με

$\delta(\boldsymbol{x},\mu;\boldsymbol{\Sigma}) = (\boldsymbol{x}-\mu)^{\boldsymbol{T}}\boldsymbol{\Sigma}^{-\mathbf{1}}(\boldsymbol{x}-\mu)$**,** να είναι όπως είδαμε η απόσταση Mahalanobis.

Λαμβάνουμε τότε τις πυκνότητες των έγκυρων και μη χρηστών ($f_{gen}(x)$ και $f_{imp}(x)$ αντιστοίχως) ως ένα μείγμα *Student t* κατανομών:

$$\hat{f}_{gen}(\boldsymbol{x}) = \sum_{i=1}^{G_{gen}}\pi_{gen,i}\varphi^G(\boldsymbol{x};\boldsymbol{\mu}_{gen,i},\boldsymbol{\Sigma}_{gen,i},\boldsymbol{v}_{gen,i})$$

και

$$\hat{f}_{imp}(\boldsymbol{x}) = \sum_{i=1}^{G_{imp}}\pi_{imp,i}\varphi^G(\boldsymbol{x};\boldsymbol{\mu}_{imp,i},\boldsymbol{\Sigma}_{imp,i},\boldsymbol{v}_{imp,i})$$

με $\pi_{gen,i}$ (και $\pi_{imp,i}$) να είναι το βάρος το οποίο ανατίθεται στον i-οστό παράγοντα της $\hat{f}_{gen}(\boldsymbol{x})$ (και $\hat{f}_{imp}(\boldsymbol{x})$ αντιστοίχως) για το οποίο ισχύει ότι:

$$\sum_{i=1}^{G_{gen}} \pi_{gen,i} = 1\,,$$

$$\sum_{i=1}^{G_{imp}} \pi_{imp,i} = 1$$

και G_{gen}, G_{imp} να είναι το πλήθος των παραγόντων που χρησιμοποιούνται για τη μοντελοποίηση των πυκνοτήτων των match scores των έγκυρων και μη χρηστών αντιστοίχως.

Για τον προσδιορισμό των τιμών των G_{gen} και G_{imp} καθώς και των παραμέτρων (διάνυσμα μέσων τιμών, πίνακας συνδιακυμάνσεων και διάνυσμα βαθμών ελευθερίας) των παραγόντων του προτεινόμενου μοντέλου μίξης χρησιμοποιείται ο ΕΜ αλγόριθμος, όπως αυτός παρουσιάστηκε στην παράγραφο **Error! Reference source not found.**. Ο αλγόριθμος αυτός γενικώς θεωρείται αρκετά εύρωστος στην αρχικοποίηση των τιμών των παραμέτρων και χειρίζεται διακριτούς παράγοντες στην κατανομή των match scores μοντελοποιώντας τα διακριτά match scores ως έναν παράγοντα μίξης με πολύ μικρή διακύμανση.

Επιστρέφοντας στο θεώρημα Neyman-Pearson ορίζουμε όπως και στο [103] τον λόγο πιθανοτήτων:

$$LR(x) = \hat{f}_{gen}(\boldsymbol{x}) / \hat{f}_{imp}(\boldsymbol{x})$$

και τον ακόλουθο κανόνα απόφασης:

Αν $LR(x) \geq \eta$, τότε ανέθεσε το x στην κλάση έγκυρων χρηστών, αλλιώς ανέθεσέ το στην κλάση των μη έγκυρων χρηστών,

με το η να αποτελεί το κατώφλι της απόφασης του κανόνα για δεδομένη τιμή του προς επίτευξη FAR.

6.4 *Αξιολόγηση Προτεινόμενου Μηχανισμού*

Η αποτελεσματικότητα του προτεινόμενου καινοτόμου μηχανισμού συνδυασμού των επι μέρους αποτελεσμάτων βιομετρικών συγκριτών βασίστηκε στη βάση NIST-BSSR1 [110] η οποία περιέχει ένα σύνολο από αληθινά match scores από δύο συστήματα αναγνώρισης προσώπου τα οποία βασίζονται στην εμπρόσθια όψη του προσώπου και ένα σύστημα αναγνώρισης δαχτυλικών αποτυπωμάτων το οποίο χρησιμοποιεί δαχτυλικά αποτυπώματα αριστερού και δεξιού δείκτη.

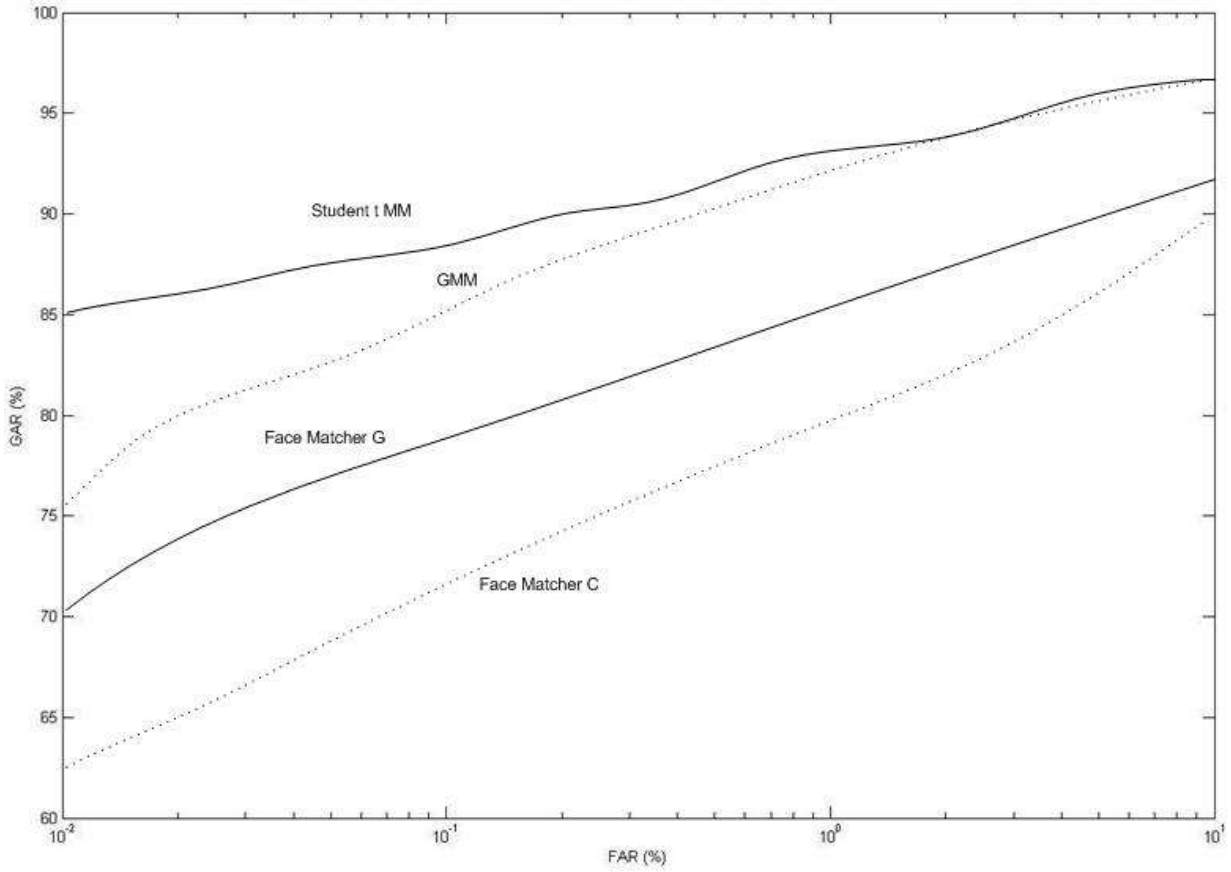
Στον ακόλουθο πίνακα συνοψίζονται τα περιεχόμενα της βάσης αυτής.

Δεδομένα Συνδυασμού	**Πλήθος Συγκριτών**	**Πλήθος Δειγμάτων**
Πρόσωπο με Πρόσωπο	2	3000
Δαχτυλικό Αποτύπωμα με Δαχτυλικό Αποτύπωμα	2	6000
Πρόσωπο με Δαχτυλικό Αποτύπωμα	Προσώπου : 2 Δαχτυλ. Αποτυπωμάτων : 4	517

Πίνακας 2 Περιεχόμενα βάσης δοκιμαστικών δεδομένων

Για κάθε πείραμα το οποίο πραγματοποιήσαμε λάβαμε τυχαία 800 έγκυρους και 800 μη έγκυρους χρήστες από τα δεδομένα της βάσης για την εκπαίδευση του συστήματος. Συνολικά έλαβαν χώρα 50 πειράματα για κάθε σημείο των ακόλουθων γραφικών παραστάσεων στις οποίες και χρησιμοποιήθηκαν οι μέσες τιμές αυτών, ενώ το διάστημα εμπιστοσύνης είναι 95%. Στις γραφικές παραστάσεις αυτές παρουσιάζεται ο GAR συναρτήσει του FAR σε ποσοστιαίες

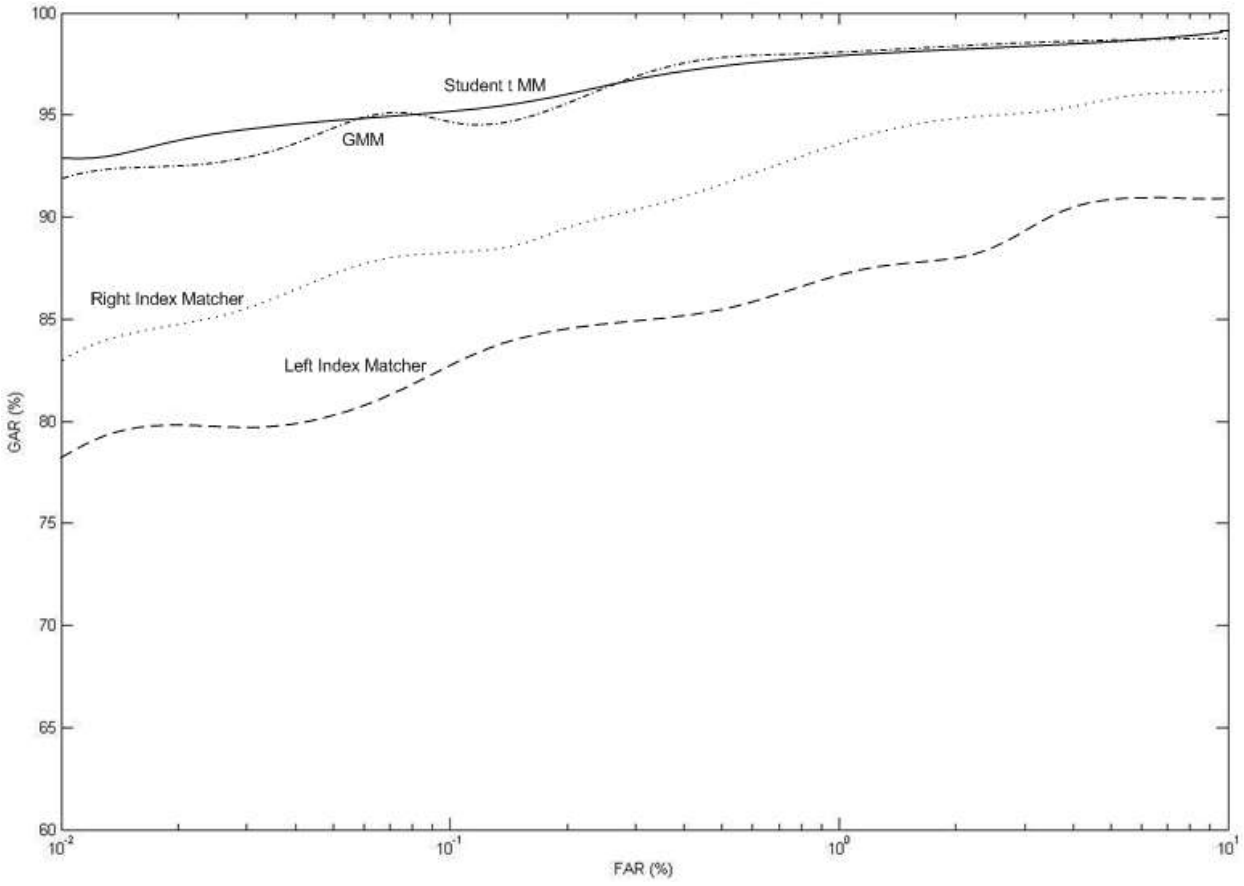
μονάδες (καμπύλη ROC), με τον άξονα του FAR να αναπαριστάται σε λογαριθμική κλίμακα. Στην Εικόνα 14 παρατίθεται σύγκριση του προτεινόμενου αλγόριθμου με τον GMM-based likelihood ratio fusion rule καθώς και τους δύο συγκριτές τους οποίους συνδυάζουν οι εν λόγω μέθοδοι και αφορά στα δεδομένα των δύο συστημάτων αναγνώρισης.



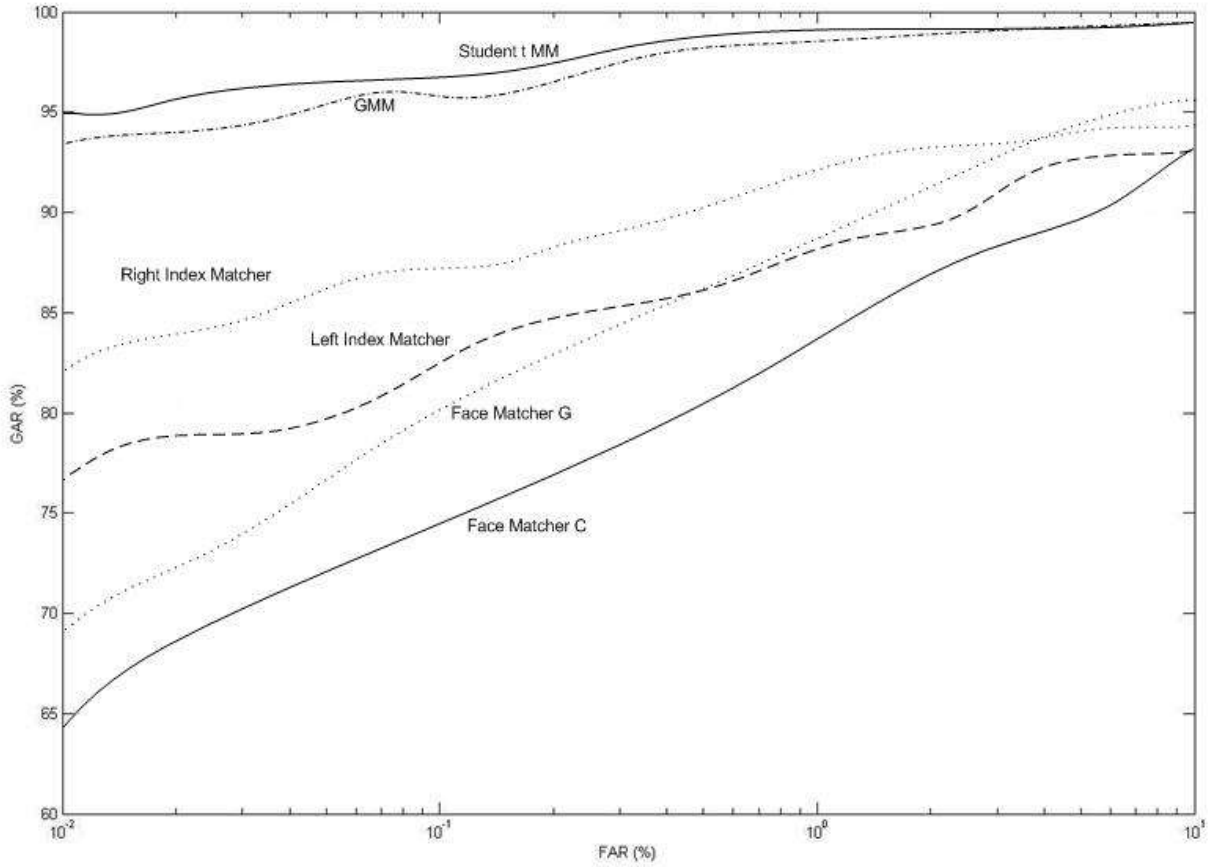
Εικόνα 14 Σύγκριση προτεινόμενου αλγόριθμου με GMM-based likelihood ratio fusion rule και τους δύο συγκριτές face matcher C και face matcher G

Όπως μπορούμε να παρατηρήσουμε και από την παραπάνω γραφική παράσταση η προτεινόμενη μέθοδος δίνει σημαντικά καλύτερες τιμές από τον κάθε face matcher των δύο συστημάτων αναγνώρισης προσώπου ξεχωριστά, ενώ υπερτερεί επίσης της μεθόδου GMM-based likelihood ratio fusion rule.

Ομοίως στις ακόλουθες γραφικές παραστάσεις εικονίζονται ο GAR συναρτήσει του FAR για το σύστημα αναγνώρισης δαχτυλικών αποτυπωμάτων το οποίο βασίζεται στα δαχτυλικά αποτυπώματα από τον αριστερό και τον δεξιό δείκτη των χεριών και για όλα τα συστήματα αναγνώρισης προσώπου και αναγνώρισης δαχτυλικών αποτυπωμάτων. Όπως μπορούμε να παρατηρήσουμε ο συνδυασμός των αποτελεσμάτων των συστημάτων σε όλες τις περιπτώσεις δίνει σημαντικά καλύτερα αποτελέσματα σε σχέση με το κάθε μονοτροπικό σύστημα ξεχωριστά. Επιπροσθέτως, ο προτεινόμενος μηχανισμός για πιθανοτικό συνδυασμό των αποτελεσμάτων βασισμένο σε Student t mixture models προκύπτει αποτελεσματικότερος από τα Gaussian Mixture Models στο μεγαλύτερο μέρος των πειραμάτων.



Εικόνα 15 Σύγκριση προτεινόμενου αλγόριθμου με GMM-based likelihood ratio fusion rule και τους δύο συγκριτές Right Index Matcher και Left Index Matcher



Εικόνα 16 Σύγκριση προτεινόμενου αλγόριθμου με συνδυασμό των βαθμολογιών όλων των συστημάτων με GMM-based likelihood ratio fusion rule, και τα επιμέρους αποτελέσματα των συστημάτων

6.5 *Συμπεράσματα*

Στο κεφάλαιο αυτό παρουσιάστηκαν δύο καινοτόμοι μέθοδοι οι οποίες εστιάζουν στη βελτίωση της αποτελεσματικότητας των βιομετρικών συστημάτων. Η προτεινόμενη αρχιτεκτονική αφορά στην εισαγωγή νέων δομικών στοιχείων τα οποία ενσωματώνονται στην υπάρχουσα αρχιτεκτονική των βιομετρικών συστημάτων προκειμένου να επιτρέψουν την εκμετάλλευση του ανεκμετάλλευτου πληροφοριακού πλούτου τόσο των βιομετρικών δεδομένων όσο και των εφαρμοζόμενων αλγόριθμων με στόχο τη βελτίωση της αποτελεσματικότητας των βιομετρικών συστημάτων με όρους αύξησης των ορθώς θετικών και αρνητικών και μείωσης των εσφαλμένως αρνητικών και θετικών. Παρουσιάστηκε επίσης ένας καινοτόμος μηχανισμός μίξης σε επίπεδο βαθμολογιών (score-level fusion) ο οποίος βασίζεται σε μοντέλα μίξης Student t κατανομών. Ο μηχανισμός αυτός εφαρμόστηκε σε πραγματικά δεδομένα στη βάση NIST-BSSR1 με τα προκύπτοντα αποτελέσματα να είναι σημαντικά βελτιωμένα τόσο έναντι των μονοτροπικών συστημάτων όσο και έναντι του μηχανισμού μίξης αποτελεσμάτων με GMM.

7 *Καινοτόμοι Μηχανισμοί Δημιουργίας και Διατήρησης Αντιγράφων σε Κατανεμημένο Περιβάλλον*

Στο κεφάλαιο αυτό, βασιζόμενοι στο γεγονός ότι ένας από τους πιο σημαντικούς πόρους σε ένα βιομετρικό σύστημα είναι τα ίδια τα δεδομένα, περιγράφονται προτεινόμενοι καινοτόμοι μηχανισμοί δημιουργίας και διατήρησης αντιγράφων σε κατανεμημένο περιβάλλον καθώς και η αξιολόγηση αυτών.

7.1 *Προσδιορισμός του Προβλήματος*

Όπως έχουμε ήδη αναφέρει, η κεντρικοποιημένη αποθήκευση των βιομετρικών δεδομένων γενικώς αποθαρρύνεται με κυριότερους λόγους αυτού να αποτελούν τόσο η διασφάλιση της ασφάλειας των βιομετρικών συστημάτων όσο και η ίδια η αποδοτικότητα και η αξιοπιστία του συστήματος. Ιδιαιτέρως στην περίπτωση βιομετρικών συστημάτων πραγματικού χρόνου και

κυρίως αυτών με κύριο στόχο είτε την ενίσχυση της ασφάλειας είτε το εμπορικό κέρδος, η ανοχή σε σφάλματα αλλά και η διαθεσιμότητα και η αποκρισιμότητα των συστημάτων αυτών αποτελούν βασικές παραμέτρους ποιότητας οι οποίες πρέπει να λαμβάνονται σοβαρά υπ'όψιν κατά τον σχεδιασμό και την υλοποίηση αυτών.

Η *διαχείριση αντιγράφων δεδομένων* (*data replication management*) αποτελεί μια προσέγγιση υψίστης σημασίας για συστήματα – και ιδιαιτέρως μεσαίας και μεγάλης κλίμακας – τα οποία αποθηκεύουν και παρέχουν δεδομένα. Το ζήτημα της διαχείρισης αντιγράφων δεδομένων αφορά στη δημιουργία και στη διατήρηση πολλαπλών πανομοιότυπων στιγμιοτύπων δεδομένων σε διάφορες τοποθεσίες. Σε προβλήματα πραγματικού κόσμου, διάφοροι περιορισμοί υφίστανται οι οποίοι αφορούν στον διαθέσιμο αποθηκευτικό χώρο, στους υπολογιστικούς πόρους, στο εύρος του δικτύου, στο κόστος διατήρησης αντιγράφων και πρόσβασης σε αυτά μεταξύ των άλλων των οποίων ο χειρισμός πρέπει να είναι κατάλληλος προκειμένου να επιτευχθεί η απαιτούμενη Ποιότητα Υπηρεσίας (Quality of Service – QoS).

Τα πλεονεκτήματα ενός βιομετρικού συστήματος το οποίο υποστηρίζει τη δημιουργία και διατήρηση αντιγράφων δεδομένων είναι πολλά. Η μη κεντρικοποιημένη αποθήκευση των βιομετρικών δεδομένων αλλά και η ύπαρξη «πολλών» αντιγράφων αυτών αυξάνει τα επίπεδα *ασφαλείας* του συστήματος καθώς επιτυχής επίθεση σε έναν αποθηκευτικό κόμβο του συστήματος θα διακυβεύσει μόνο τα δεδομένα τα οποία είναι αποθηκευμένα στον κόμβο αυτόν. Επιπροσθέτως, η *αξιοπιστία* του βιομετρικού συστήματος αυξάνεται, καθώς η διασπορά των βιομετρικών δεδομένων αναφοράς σε πολλαπλούς κόμβους εξασφαλίζει τη διαθεσιμότητα αυτών ακόμη και όταν ένας από αυτούς τους κόμβους πάψει να λειτουργεί λόγω σφάλματος σε επίπεδο υλικού ή λογισμικού, απώλειας δικτύου ή πτώσης ρεύματος. Η *απόδοση* του συστήματος επίσης βελτιώνεται σημαντικά μέσω της μείωσης των χρονικών καθυστερήσεων

πρόσβασης στα δεδομένα, καθώς μέσω των αντιγράφων των δεδομένων οι αιτήσεις πρόσβασης σε αυτά θα μπορούν να δρομολογούνται πιο έξυπνα στους λιγότερο «φορτωμένους» κόμβους.

7.2 *Σχετική Έρευνα*

Πολλές μέθοδοι και μηχανισμοί έχουν προταθεί κατά καιρούς για τη διαχείριση αντιγράφων σε κατανεμημένα περιβάλλοντα. Έτσι, οι Goel και Buyya παρουσιάζουν μια έρευνά τους σχετικά με αλγόριθμους διατήρησης αντιγράφων σε διαφορετικά συστήματα κατανεμημένης αποθήκευσης και διαχείρισης περιεχομένου συμπεριλαμβάνοντας αρχιτεκτονικές Πλέγματος [111], ενώ οι Saito και Shapiro περιγράφουν μια έρευνά τους σχετική με αισιόδοξους αλγόριθμους δημιουργίας αντιγράφων οι οποίοι επιτρέπουν στα αντίγραφα να παρεκλίνουν βραχυπρόθεσμα [112].

Οι Lei et al περιγράφουν μια κατανεμημένη στραγηγική για τον τοπικό και χρονικό καθορισμό της δημιουργίας αντιγράφων με στόχο τη βελτίωση της διαθεσιμότητας του συστήματος [113], ενώ σε μεταγενέστερη δημοσιευμένη εργασία τους [114] παρουσιάζουν τέσσερεις βελτιστοποιητές οι οποίοι σχετίζονται με τέσσερεις διαφορετικές συναρτήσεις πρόβλεψης πρόσβασης αρχείων προκειμένου να αντιμετωπίσουν το ζήτημα της διαθεσιμότητας δεδομένων θεωρώντας περιορισμένο αποθηκευτικό χώρο για τα αντίγραφα. Οι Lamehamedi et al [115] περιγράφουν ένα σύνολο από υπηρεσίες και πρωτόκολλα διαχείρισης αντιγράφων με τις αποφάσεις δημιουργίας αντιγράφων να βασίζονται σε ένα μοντέλο κόστους το οποίο αποτιμά το κόστος πρόσβασης δεδομένων και τα κέρδη απόδοσης από τη δημιουργία κάθε αντιγράφου, ενώ τα αντίγραφα οργανώνονται σε έναν συνδυασμό ιεραρχικών και επίπεδων τοπολογιών για την ελαχιστοποίηση των επικοινωνιακών καθυστερήσεων. Οι Cameron et al [116] παρουσιάζουν και αξιολογούν μια στρατηγική δημιουργίας αντιγράφων η οποία είναι βασισμένη σε ένα οικονομικό

μοντέλο το οποίο στοχεύει στη βελτιστοποίηση τόσο της επιλογής των αντιγράφων όσο και στη δυναμική δημιουργία αντιγράφων.

Το θέμα της διαχείρισης αντιγράφων δεδομένων έχει αποκτήσει έντονο ενδιαφέρον τα τελευταία χρόνια και στον χώρο των τεχνολογιών Πλέγματος, με κύριο στόχο τη διαθεσιμότητα των δεδομένων και την απόδοση πρόσβασης σε αυτά. Συγκεκριμένα, οι Chang και Chen παρουσιάζουν μια προσέγγιση για τμηματοποιημένα αντίγραφα προκειμένου να σώσουν αποθηκευτικό χώρο [117]. Οι Lin et al [118] παρουσιάζουν έναν replication broker ο οποίος μειώνει τις καθυστερήσεις των μηχανισμών δημιουργίας αντιγράφων μέσω της επέκτασης του resource broker ώστε να λάβει υπ'όψιν τακτικές σχετικά με τη μεταφορά δεδομένων και την αντιγραφή αυτών.

Οι Guy et al [119] παρουσιάζουν την αρχιτεκτονική και τον σχεδιασμό ενός συστήματος διαχείρισης αντιγράφων, ονόματι «*Reptor*», το οποίο εστιάζει σε ζητήματα θέσης και χρόνου ζωής των αντιγράφων. Αποτέλεσμα μιας άλλης ερευνητικής εργασίας ήταν η εισαγωγή τεχνικών πολυεκπομπής (multicast) με έναν κατάλογο αντιγράφων (replica catalog) ο οποίος αυτομάτως ανιχνεύει ταυτόχρονες προσπάθειες πρόσβασης σε ένα αντίγραφο από πολλαπλούς κόμβους [120].

Οι προτεινόμενοι μηχανισμοί δημιουργίας και διατήρησης αντιγράφων οι οποίοι θα παρουσιαστούν στη συνέχεια διαφέρουν από τους παρουσιασθέντες στην παρούσα παράγραφο ως προς το γεγονός ότι οι πλείστοι από τους τελευταίους στοχεύει στην αντιμετώπιση συγκεκριμένων ζητημάτων, όπως είναι η τοποθέτηση αντιγράφων, αρκετοί εξ αυτών είναι στατικοί, ενώ ακόμη και αυτοί οι οποίοι είναι δυναμικοί δε βασίζονται σε παραμέτρους διασφάλισης Ποιότητας Υπηρεσίας όπως είναι το εύρος του δικτύου, το κόστος αποθήκευσης και διατήρησης του αντιγράφου και η τοπικότητα των αιτημάτων.

Στις επόμενες παραγράφους θα παρουσιάσουμε ένα σύνολο από αλγόριθμους δημιουργίας, τοποθέτησης, αναδιανομής και διαγραφής αντιγράφων, οι οποίοι μπορούν και *διαλειτουργούν,* παρέχοντας έτσι μια ολοκληρωμένη προσέγγιση διαχείρισης αντιγράφων. Συγκεκριμένα, η παρουσιαζόμενη λύση δυναμικής διαχείρισης αντιγράφων εστιάζει σε αποφάσεις σχετικά με:

- το *πλήθος* των *αντιγράφων* τα οποία θα *δημιουργηθούν*,
- την *τοποθεσία* στην οποία θα *αποθηκευτούν* τα αντίγραφα,
- τη *χρονική στιγμή* κατά την οποία ένα αντίγραφο θα *διαγραφεί*,
- την *τοποθεσία* στην οποία θα *επανατοποθετηθεί* ένα αντίγραφο

με κύριο στόχο την *αύξηση της διαθεσιμότητας των δεδομένων*, τη *βελτίωση της απόδοσης του συστήματος* μέσω της *εξισορρόπησης των αιτήσεων πρόσβασης* αλλά και της *βελτίωσης της χρήσης του εύρους του δικτύου*.

Στο σημείο αυτό πρέπει να τονίσουμε ότι αν και στο γενικότερο ερευνητικό πλαίσιο της διαχείρισης αντιγράφων πολύ σημαντικός μηχανισμός είναι ο συγχρονισμός των δεδομένων, σε ένα βιομετρικό σύστημα η ενημέρωση των βιομετρικών δεδομένων και συνεπώς η αλλαγή αυτών δεν αποτελεί σύνηθες φαινόμενο – λόγω της μοναδικότητας αυτών αλλά και της διαχρονικότητάς τους – και για τον λόγο αυτόν δε μελετήθηκε στα πλαίσια της παρουσιαζόμενης έρευνας.

7.3 *Προτεινόμενοι καινοτόμοι μηχανισμοί διαχείρισης αντιγράφων σε κατανεμημένο περιβάλλον*

Οι αλγόριθμοι οι οποίοι παρουσιάζονται στις επόμενες παραγράφους έχουν υλοποιηθεί για ένα σύστημα με κεντρικοποιημένη αρχιτεκτονική, η οποία περιλαμβάνει ένα μεγάλο πλήθος

αποθηκευτικών κόμβων και έναν κόμβο διαχείρισης αντιγράφων στον οποίο διατηρείται ένας κατάλογος αντιγράφων και ο οποίος διαχειρίζεται την πρόσβαση στα αντίγραφα καθώς και τον κύκλο ζωής τους. Όπως θα δούμε και στις παρουσιαζόμενες λύσεις, οι μηχανισμοί αυτοί φέρουν εγγενείς ομοιότητες και απαιτούν την επίλυση ενός υπολογιστικά απαιτητικού προβλήματος, το οποίο και αντιμετωπίζουμε στην παράγραφο 7.3.5.

7.3.1 Δημιουργία Αντιγράφων

Προσεγγίζοντας το ζήτημα της δημιουργίας αντιγράφων από την πλευρά των περιορισμών, τόσο οι δικτυακοί όσο και οι αποθηκευτικοί πόροι οι οποίοι καταναλώνονται για την αποθήκευση και τη διαχείριση των αντιγράφων δεν είναι ατέρμονοι. Έτσι, η δημιουργία αντιγράφων δεδομένων πρέπει να λαμβάνει χώρα όταν πληρούνται κάποιες προϋποθέσεις και όχι τυχαία. Για τον λόγο αυτό εισάγουμε μια παράμετρο στο πρόβλημα, τον *συνολικά διαθέσιμο χώρο* ο οποίος μπορεί να διατεθεί για τη δημιουργία νέων αντιγράφων. Έστω ότι η παράμετρος αυτή συμβολίζεται με *S*.

7.3.1.1 Ορισμός του παράγοντα «δημοτικότητας»

Όπως έχουμε αναφέρει ο πιο σημαντικός πόρος ενός βιομετρικού συστήματος είναι τα ίδια τα βιομετρικά δεδομένα, με τις απαιτήσεις για *διαθεσιμότητα* και *σταθερότητα* να ποικίλλουν αναλόγως με την εφαρμογή των βιομετρικών συστημάτων. Στο σημείο αυτό θα πρέπει να τονίσουμε ότι δεδομένου του όγκου των βιομετρικών δεδομένων όσον αφορά σε εφαρμογές μικρής και μεγάλης κλίμακας είναι πολύ πιθανό ο πάροχος του βιομετρικού συστήματος να «νοικιάζει» αποθηκευτικούς πόρους από κάποιον άλλο πάροχο/παρόχους. Από τα βιομετρικά δεδομένα αυτά, τα οποία κυρίως αφορούν τα πρότυπα αναφοράς, κάποια σύνολα δεδομένων είναι πιο «*δημοφιλή*» και, συνεπώς, η δημιουργία αντιγράφων αυτών είναι πιο σημαντική,

έχοντας ως στόχο την εξισορρόπηση του φόρτου στο σύστημα, ενώ κάθε εφαρμογή η οποία αιτείται πρόσβαση σε αυτά μπορεί να επαναδρομολογηθεί σε ένα αντίγραφο το οποίο βρίσκεται σε «κοντινή απόσταση» στον αιτόντα κόμβο, επιτρέποντας γρηγορότερη πρόσβαση.

Προκειμένου να λάβουμε υπ'όψιν τη «δημοτικότητα» των δεδομένων, ο κόμβος διαχείρισης αντιγράφων κρατά αρχεία των αιτήσεων πρόσβασης, συμπεριλαμβανομένης της ακριβούς ημερομηνίας και ώρας κατά την οποία έλαβε χώρα το αίτημα καθώς και τη γεωγραφική τοποθεσία από την οποία έγινε. Η πληροφορία αυτή λαμβάνεται υπ'όψιν για τον υπολογισμό της *μετρικής «δημοτικότητας»* (*popularity metric*) για κάθε δεδομένο. Τα δεδομένα τα οποία συλλέγονται αφορούν σε ένα συγκεκριμένο χρονικό παράθυρο το οποίο καθορίζεται από τις απαιτήσεις της εφαρμογής.

Μετρώντας το *πλήθος των αιτήσεων ανάκτησης* στο ορισμένο χρονικό παράθυρο, λαμβάνουμε μια κατανομή επί των δεδομένων του συστήματος $r(i)$, $i = 1,\ldots,k$, με k να είναι τα δεδομένα στο σύστημα. Έτσι, το συνολικό πλήθος αιτήσεων R για ανάκτηση δεδομένων σε ένα δεδομένο χρονικό παράθυρο δίνεται από τη σχέση :

$$R = \sum_{i=1}^{k} r(i)$$

Εκτός όμως από το πλήθος των αιτήσεων, και το κόστος αποθήκευσης και διατήρησης των δεδομένων πρέπει να λαμβάνεται υπ'όψιν κατά τη διαχείριση των αντιγράφων. Στον επιχειρηματικό χώρο είναι αρκετά σύνηθες να υφίστανται διάφορα «πακέτα» υπηρεσίας τα οποία τιμολογούνται διαφορετικά αναλόγως με το επίπεδο της προσφερόμενης υπηρεσίας.

Προκειμένου να συμπεριλάβουμε και τις συνθήκες αυτές στο σύστημά μας, εισαγάγαμε σε πρώτη φάση τον *παράγοντα Ποιότητας Υπηρεσίας*. Ο παράγοντας αυτός λαμβάνει τιμές μεταξύ 0 και 1, με τη σχέση μεταξύ του παράγοντα αυτού και της παρεχόμενης ποιότητας υπηρεσίας να

είναι ανάλογη. Στα πλαίσια της μελέτης μας ο παράγοντας αυτός ενσωματώνει τις απαιτήσεις για διαθεσιμότητα και απόδοση. Ο παράγοντας αυτός προσαρμόζει την κατανομή των απαιτήσεων *r(i)*QoSfactor(i), i = 1,…, k,* με τον *QoSfactor(i)* να αφορά στην ποιότητα υπηρεσίας την οποία ο πάροχος του βιομετρικού συστήματος απαιτεί για το i-οστό σύνολο δεδομένων. Κανονικοποιώντας την κατανομή των QoS-aware αιτήσεων *f(i)* καταλήγουμε στην ακόλουθη σχέση:

$$f(i) = \frac{r(i) \cdot QoSfactor\ (i)}{\sum_{i=1}^{k}\{r(i) \cdot QoSfactor\ (i)\}} \quad (1)$$

για την οποία προφανώς ισχύει ότι:

$$\sum_{i=1}^{k} f(i) = 1$$

7.3.1.2 Κατανομή του διαθέσιμου αποθηκευτικού χώρου

Ο χώρος ο οποίος χρησιμοποιείται για τη δημιουργία των αντιγράφων για το i-οστό σύνολο δεδομένων θα πρέπει να είναι ανάλογο του $f(i)$, δηλαδή:

$$\tilde{s}(i) = f(i) \cdot S \qquad (2)$$

Μια προφανής ταυτότητα είναι η ακόλουθη:

$$\sum_{i=1}^{k} \tilde{s}(i) = \sum_{i=1}^{k}\{f(i) \cdot S\} = S \cdot \sum_{i=1}^{k} f(i) = S \qquad (3)$$

Από τις σχέσεις (1) και (2) μπορεί εύκολα να εξαχθεί ότι ο όρος $\tilde{s}(i)$ είναι γενικώς πραγματικός αριθμός. Όμως, ο αποθηκευτικός χώρος θα πρέπει να εκφράζεται ως ακέραιο πολλαπλάσιο της μικρότερης μονάδας αποθήκευσης (συνήθως χρησιμοποιείται το byte). Καθώς ο *S* είναι ένας πολύ μεγάλος αριθμός σε σχέση με το *k*, μπορούμε να στρογγυλοποιήσουμε προς τα κάτω τον όρο $\tilde{s}(i)$ χωρίς τον κίνδυνο να μείνει μεγάλο μέρος του *S* ανεκμετάλλευτο.

$$\mathrm{s(i)} = \lfloor \tilde{\mathrm{s}}(\mathrm{i}) \rfloor = \lfloor \mathrm{f(i)} \cdot \mathrm{S} \rfloor \qquad (4)$$

Με βάση τις σχέσεις (3) και (4) ο χώρος ο οποίος μένει ανεκμετάλλευτος είναι ίσος με:

$$S - \sum_{i=1}^{k} s(i) = \sum_{i=1}^{k} \tilde{s}(i) - \sum_{i=1}^{k} s(i) = \sum_{i=1}^{k} \{\tilde{s}(i) - s(i)\} = \sum_{i=1}^{k} \{\tilde{s}(i) - \lfloor \tilde{s}(i) \rfloor\} < k$$

Η επίλυση του προβλήματος της δημιουργίας αντιγράφων τώρα έγκειται στην εύρεση του πλήθους των αντιγράφων τα οποία πρέπει να δημιουργηθούν για κάθε σύνολο δεδομένων. Έστω ότι $\tilde{n}(i)$ είναι το πλήθος των αντιγράφων τα οποία πρέπει να δημιουργηθούν για το i-οστό σύνολο δεδομένων, με τη μαθηματική διατύπωση αυτού να είναι:

$$\tilde{n}(i) \cdot size(i) = s(i) \Rightarrow \tilde{n}(i) = s(i)/size(i) \quad (5)$$

7.3.1.3 Αλγόριθμος γραμμικής πολυπλοκότητας

Ο πιο απλός τρόπος υπολογισμού του *n(i), i = 1,..., k* είναι η στρογγυλοποίηση προς τα κάτω:

$$n(i) = \lfloor \tilde{n}(i) \rfloor = \lfloor s(i)/size(i) \rfloor \quad (6)$$

Ο αλγόριθμος αυτός είναι αρκετά αποδοτικός, καθώς είναι πολυπλοκότητας *O(k)*. Όμως, μπορεί να αφήσει μεγάλο μέρος του *S* ανεκμετάλλευτο. Πιο συγκεκριμένα, το άνω όριο του αποθηκευτικού χώρου ο οποίος μπορεί να μείνει ανεκμετάλλευτος εξαιτίας της στρογγυλοποίησης είναι:

$$\sum_{i=1}^{k} size(i)$$

Πράγματι, από τη σχέση (6) και τη γνωστή ανισότητα $\lfloor x \rfloor \leq x \leq \lfloor x \rfloor + 1$, αυτομάτως προκύπτει ότι :

$$n(i) \leq s(i)/size(i) < n(i) + 1 \Leftrightarrow n(i) \cdot size(i) \leq s(i) \leq n(i) \cdot size(i) + size(i)$$

και αθροίζοντας για όλα τα k σύνολα δεδομένων :

$$\sum_{i=1}^{k} n(i) \cdot size(i) \leq \sum_{i=1}^{k} s(i) \leq \sum_{i=1}^{k} n(i) \cdot size(i) + \sum_{i=1}^{k} size(i) \Rightarrow$$

$$0 \leq \sum_{i=1}^{k} s(i) - \sum_{i=1}^{k} n(i) \cdot size(i) < \sum_{i=1}^{k} size(i)$$

7.3.1.4 Αλγόριθμος O(k²)πολυπλοκότητας

Ο προτεινόμενος αλγόριθμος σε αυτήν την παράγραφο υπερέχει του προηγούμενου αλγόριθμου ως προς το γεγονός ότι ο χώρος ο οποίος μένει ανεκμετάλλευτος λόγω της στρογγυλοποίησης είναι μικρότερος από το μέγεθος του μικρότερου συνόλου δεδομένων.

Τα βήματα του αλγόριθμου αυτού είναι:

- Εύρεση του συνόλου δεδομένων με το μεγαλύτερο μέγεθος (έστω ότι είναι το j-οστό)

$$size(j) = \max_{i=1,\dots,k} \{size(i)\}$$

- Υπολογισμός του πλήθους των αντιγράφων τα οποία θα δημιουργηθούν μέσω της σχέσης:

$$n(j) = \lfloor s(j)/size(j) \rfloor$$

- Εξαιτίας της στρογγυλοποίησης ένα μέρος του χώρου *s(j)* παραμένει αχρησιμοποίητο, το οποίο ισούται με:

$$leftSpace = s(j) - n(j) \cdot size(j)$$

και το οποίο θα μοιραστεί στα υπόλοιπα σύνολα δεδομένων, (k-1) στο πλήθος.

Δεδομένου ότι η διαίρεση $leftSpace/(k-1)$ δε δίνει γενικώς ακέραιο αριθμό, ο αλγόριθμος διαίρεσης του $leftSpace$ σε (k-1) μέρη όσο πιο ομοιόμορφα είναι δυνατό είναι ο ακόλουθος:

- Έστω ότι *leftSpace = a · (k-1)+b* είναι η Ευκλείδια διαίρεση του απομείνοντος χώρου $leftSpace$ με το (k-1), με το *α* να είναι το πηλίκο της διαίρεσης και *b* το υπόλοιπο.
- Τότε, *b* μέρη θα είναι ίσα με *(α+1)* και *(k-1-b)* μέρη θα είναι ίσα με *α*.

$$leftSpace = \underbrace{(a+1) + \ldots + (a+1)}_{b} + \underbrace{a + \cdots + a}_{k-1-b}$$

Πράγματι,

$$b \cdot (a+1) + (k-1-b) \cdot a = a \cdot b + b + a \cdot (k-1) - a \cdot b =$$

$$= b + a \cdot (k-1) = leftSpace$$

Έτσι, $s'(i) = s(i) + (a+1)$ για b σύνολα δεδομένων and $s'(i) = s(i) + a$ για (k - *1* - b) σύνολα δεδομένων.

Η ίδια διαδικασία ακολουθείται για τα απομείναντα (k-1) σύνολα δεδομένων χρησιμοποιώντας τις νέες τιμές $s'(i)$ όπως αυτές υπολογίζονται από το προηγούμενο βήμα.

Ο αλγόριθμος μπορεί να παραλαχθεί ελαφρώς ώστε να αυξηθεί η απόδοσή του. Συγκεκριμένα, με την ταξινόμηση των αρχείων κατά φθίνουσα σειρά μεγεθών, το πρώτο βήμα του αλγόριθμου μπορεί να παραληφθεί. Η πολυπλοκότητα της ταξινόμησης είναι *O(k * log k)*. Στον Πίνακας 3 παρουσιάζεται η πρώτη επανάληψη του τροποποιημένου αλγόριθμου. Θα πρέπει να τονιστεί ότι οι δύο αλγόριθμοι οι οποίοι παρουσιάστηκαν δεν παρέχουν γενικώς τα ίδια αποτελέσματα, καθώς ο πρώτος οδηγεί πιο γρήγορα σε αποτέλεσμα αφήνοντας όμως μεγαλύτερο μέρος του διαθέσιμου χώρου αχρησιμοποίητο, ενώ ο δεύτερος αντισταθμίζει τις απαιτήσεις σε υπολογιστική ισχύ με την καλύτερη συμμόρφωση στους περιορισμούς του προβλήματος και βελτιωμένη εκμετάλλευση των διαθέσιμων αποθηκευτικών πόρων.

i	$size(i)$	$s(i)$	$n(i)$	$s'(i)$
i_1	$size(i_1)$	$s(i_1)$	$n(i_1)$	
i_2	$size(i_2)$	$s(i_2)$		$s(i_2)+(a+1)$
$\vdots$	$\vdots$	$\vdots$		$\vdots$
i_{b+1}	$size(i_{b+1})$	$s(i_{b+1})$		$s(i_{b+1})+(a+1)$
i_{b+2}	$size(i_{b+2})$	$s(i_{b+2})$		$s(i_{b+2})+a$
$\vdots$	$\vdots$	$\vdots$		$\vdots$
i_k	$size(i_k)$	$s(i_k)$		$s(i_k)+a$

Πίνακας 3 Η πρώτη επανάληψη του αλγόριθμου δημιουργίας αντιγράφων

7.3.2 Τοποθέτηση Αντιγράφων

Το πρόβλημα της τοποθέτησης των δεδομένων μπορεί να επιλυθεί για κάθε σύνολο δεδομένων ξεχωριστά. Συνεπώς, είναι παραλληλίσιμο σε μεγάλο βαθμό και συνακολούθως μπορεί να επιταχυνθεί σημαντικά η επίλυσή του.

7.3.2.1 Ορισμός μιας μετρικής απόστασης

Κατά τη διαδικασία επιλογής των καλύτερων αποθηκευτικών κόμβων για την αποθήκευση των νέων αντιγράφων, ένα από τα πιο σημαντικά κριτήρια είναι η *τοπικότητα (locality)*. Ένας αποθηκευτικός χώρος θα είναι προτιμητέος αν είναι σε «κοντινή» θέση σε σχέση με τις συχνές αιτήσεις πρόσβασης προς τα δεδομένα. Έτσι, ορίζουμε μια *μετρική απόστασης* η οποία συμπεριλαμβάνει παράγοντες όπως είναι η γεωγραφική απόσταση, η τοπολογία του δικτύου και το εύρος του δικτύου, μεταξύ των άλλων. Κατά τον υπολογισμό αυτής της μετρικής πρέπει επίσης να λαμβάνονται υπ'όψιν τόσο η καθυστέρηση πρόσβασης όσο και η διαθεσιμότητα, καθώς αποτελούν δύο πολύ σημαντικές παραμέτρους Ποιότητας Υπηρεσίας για το βιομετρικό σύστημα οι οποίες είναι άρρηκτα συνδεδεμένες με τον αναμενόμενο φόρτο εργασίας και την αξιοπιστία των αποθηκευτικών κόμβων.

	1	2	...	s	b_i
1	$d(1,1)$	$d(1,2)$	...	$d(1,s)$	b_1
2	$d(2,1)$	$d(2,2)$	...	$d(2,s)$	b_2
3	$d(3,1)$	$d(3,2)$	...	$d(3,s)$	b_3
⋮	⋮	⋮	⋱	⋮	⋮
r	$d(r,1)$	$d(r,2)$	...	$d(r,s)$	b_r

Πίνακας 4 Μετρική απόστασης μεταξύ αποθηκευτικού κόμβου και κόμβου αίτησης

7.3.2.2 Το πρόβλημα βελτιστοποίησης

Έστω ότι, για ένα δεδομένο σύνολο δεδομένων, *n* είναι το πλήθος των νέων αντιγράφων τα οποία θα δημιουργηθούν, *m* είναι το πλήθος των υπαρχόντων αντιγράφων, *s* είναι το πλήθος των αποθηκευτικών κόμβων και *r* είναι το πλήθος των αιτήσεων για το σύνολο δεδομένων σε ένα συγκεκριμένο χρονικό παράθυρο.

Προκειμένου να παρουσιάσουμε τη μαθηματική ανάλυση του προβλήματος θα ορίσουμε τα σύνολα *S, S_1, S_2, R* ως εξής:

Set	Description	Properties
S	the set of all the storage nodes	$\|S\| = s$
S_1	the set of the storage nodes that already contain a replica	$\|S_1\| = m$ $S_1 \subseteq S$
S_2	the set of the storage nodes that do not contain a replica yet	$\|S_2\| = t = s - m$ $S_2 \subseteq S$ $S_1 \cap S_2 = \emptyset$ $S_1 \cup S_2 = S$
R	the set of the requests made for the data set in a given time window	$\|R\| = r$
R_1	the set of the requests for which $b_i \in S_1$	$R_1 \subseteq R$
R_2	the set of the requests for which $b_i \notin S_1 \Leftrightarrow b_i \in S_2$	$R_2 \subseteq R$ $R_1 \cap R_2 = \emptyset$ $R_1 \cup R_2 = R$
$C_n^{S_2}$	the set of all the n-combinations from set S_2	$\left\|C_n^{S_2}\right\| = \binom{t}{n}$

Πίνακας 5 Σύνολα αποθηκευτικών χώρων και αιτήσεων

Είναι προφανές ότι η επίλυση του προβλήματος της τοποθέτησης των αντιγράφων ανάγεται στην επιλογή *n* αποθηκευτικών κόμβων από το σύνολο S_2, με βάση ένα κριτήριο βελτιστοποίησης. Η επιλογή *n* στοιχείων από ένα σύνολο *(s-m)* στοιχείων είναι το γνωστό πρόβλημα εύρεσης n-συνδυασμών. Για ευκολία ορίζουμε: *t=s-m*. Έτσι, το πλήθος των n-συνδυασμών (καθένας μεγέθους *n*) από ένα σύνολο με *t* στοιχεία είναι :

$$\binom{t}{n} = \frac{t!}{n!(t-n)!} \quad (7)$$

Το σύνολο το οποίο περιέχει όλους τους n-συνδυασμούς μπορεί να οριστεί ως:

$$C_n{}^{S_2} = \{C \mid C \subseteq S_2 \wedge |C| = n\} \quad (8)$$

Έστω ότι έχουμε επιλέξει έναν *n*-συνδυασμό $C_n \in C_n{}^{S_2}$ και έχουμε τοποθετήσει τα νέα αντίγραφα στους κόμβους τους οποίους περιλαμβάνει. Έτσι, το σύνολο των αποθηκευτικών

κόμβων οι οποίοι περιέχουν τώρα ένα αντίγραφο του συγκεκριμένου συνόλου δεδομένων είναι $S_1 \cup C_n$. Αν γίνει μια αίτηση για το σύνολο δεδομένων αυτών, ο πιο «κοντινός» αποθηκευτικός κόμβος πρέπει να επιλεγεί για να εξυπηρετήσει αυτήν την αίτηση. Η τελευταία στήλη του Πίνακας 5 δείχνει τον κοντινό αποθηκευτικό χώρο, με την υπόθεση ότι όλοι οι αποθηκευτικοί κόμβοι του συστήματος λαμβάνονται υπ'όψιν. Θα πρέπει να τονίσουμε ότι δεν υπάρχει απαραιτήτως ένας μοναδικός «πιο κοντινός» κόμβος και για τον λόγο αυτόν η επιλογή μεταξύ σχεδόν ισοδύναμων κόμβων μπορεί να γίνει τυχαία.

Με βάση τα ιστορικά δεδομένα των αιτήσεων συνόλων δεδομένων τα οποία περιέχονται στο σύνολο R, μπορούμε να θεωρήσουμε ότι το «κόστος» το οποίο συνεπάγεται η αίτηση $i \in R$ (c_r) είναι ανάλογο της απόστασης μεταξύ του i-οστού κόμβου αίτησης και του «πιο κοντινού» αποθηκευτικού κόμβου, θεωρώντας ότι το κόστος αποθήκευσης είναι ίσο για όλους τους κόμβους αποθήκευσης ($c_{s,j} = c_s, j \in S_1 \cup C_n$) και επομένως λαμβάνοντας υπ'οψιν μόνο τις μεταβολές της απόστασης:

$$c_r = a \cdot \min_{j \in S_1 \cup C_n}\{d(i,j)\} \quad (9)$$

Το συνολικό κόστος των αιτήσεων από το σύνολο R έτσι προκύπτει:

$$Cost\ (C_n) = a \cdot \sum_{i \in R} \min_{j \in S_1 \cup C_n}\{d(i,j)\} \quad (10)$$

ενώ λαμβάνοντας υπ'οψιν και το κόστος αποθήκευσης έχουμε ότι :

$$TotalCost\ (C_n) = a \cdot \sum_{i \in R} \min_{j \in S_1 \cup C_n}\{d(i,j)\} + n \cdot c_s$$

Συνεπώς, το κριτήριο βελτιστοποίησης για την επιλογή των στοιχείων του C_n αφορά στην ελαχιστοποίηση του συνολικού κόστους και πιο συγκεκριμένα του μεταβλητού όρου του κόστους. Έστω ότι $\hat{C}_n$ είναι ο βέλτιστος n-συνδυασμός (ο οποίος δεν είναι απαραιτήτως μοναδικός). Τότε :

$$Cost(\hat{C}_n) = \min_{C \in C_n{}^{S_2}} \{Cost(C)\} \quad (11)$$

or, equivalently,

$$\hat{C}_n = \operatorname{argmin}_{C \in C_n{}^{S_2}} \{Cost(C)\} \quad (12)$$

7.3.2.3 Φιλτράρισμα Αιτήσεων

Οι «πιο κοντινοί» αποθηκευτικοί κόμβοι $b_1,\ldots,b_r$ προσδιορίζονται στην τελευταία στήλη του Πίνακας 4. Θα πρέπει να τονιστεί ότι ο πίνακας αυτός περιέχει όλες τις αποστάσεις μεταξύ των κόμβων, ενώ μετά το πέρας της διαδικασίας μόνο μερικοί από τους κόμβους αυτούς θα περιέχουν αντίγραφα των δεδομένων ($S_1 \cup \hat{C}_n$). Μάλιστα, είναι πιθανό οι «καλύτεροι» κόμβοι στο *S* με το κριτήριο της απόστασης και συνεπώς του κόστους να μην είναι στο σύνολο $S_1 \cup \hat{C}_n$. Όμως, αν ένας από τους «πιο κοντινούς» αποθηκευτικούς κόμβους περιέχει ήδη αντίγραφο ($b_i \in S_1$), τότε είναι σίγουρο ότι είναι ο «πιο κοντινός» αποθηκευτικός κόμβος μεταξύ των αυτών στο σύνολο $S_1 \cup \hat{C}_n$. Αυτός ο ισχυρισμός αποδεικνύεται ακολούθως.

Έστω I είναι ένα πεπερασμένο σύνολο και I_1 είναι υποσύνολο αυτού, δηλαδή $I_1 \subseteq I$. Αν m = min{f(i)}, i $\in$ I και m_1 = min{f(i)}, i$\in I_1$, τότε m ≤ m_1. Πράγματι, εξ ορισμού έχουμε ότι m = min{f(i)}, i $\in$ I $\Leftrightarrow \exists\, i_0 \in I\ [m = f(i_0) \wedge \forall i \in I(m \leq f(i)]$

και

$$m_1 = \min\{f(i)\},\ i \in I_1 \Leftrightarrow \exists\, i_1 \in I_1\ [m_1 = f(i_1) \wedge \forall i \in I_1(m_1 \leq f(i)]$$

Ας υποθέσουμε ότι ισχύει ότι that m > m_1. Προφανώς $i_1 \in I_1$ συνεπάγεται ότι $i_1 \in I$, αφού $I_1 \subseteq I$. Συνεπώς, $i_1 \in I$, $m_1 = f(i_1)$ και $\forall i \in I\ (m_1 < m \leq f(i))$. Αυτό σημαίνει ότι m > m_1 = min { $f(i_1\}$, $i \in I$, το οποίο είναι άτοπο.

Με βάση τα παραπάνω μπορούμε να αποδείξουμε ότι αν argmin{d(i, j)} = $b_i \in S_1$, j ∈ S, τότε για κάθε $C_n \in C_n{}^{S_2}$ έχουμε ότι:

$$b_i = \text{argmin}_{j \in S_1 \cup C_n}\{d(i,j)\}$$

ή ισοδύναμα αν υπάρχει $b_i \in S_1$τέτοιο ώστε d(i, b_i) = min {d(i, j)} , j ∈ S, τότε για κάθε $C_n \in C_n^{S_2}$, ισχύει ότι

$$d(i, b_i) = \min_{j \in S_1 \cup C_n}\{d(i,j)\}$$

Πράγματι, έστω $m_i = d(i, b_i) = min\ \{d(i,j)\}$, $j \in S$ και $M_i = min\ \{d(i,j)\}$, $j \in S_1 \cup C_n$. Ισχύει ότι $S_1 \cup C_n \in S$ και με βάση τα παραπάνω έχουμε $m_i \leq M_i$. Αρκεί, λοιπόν, να δείξουμε ότι $M_i \leq m_i$. Εξ ορισμού $\forall j \in S_1 \cup C_n$ $(M_i \leq min\ \{d(i,j)\})$. Συνεπώς, $b_i \in S_1 \Longrightarrow b_i \in S_1 \cup C_n \Longrightarrow M_i \leq d(i, b_i)\} = m_i$.

Στον Πίνακας 5 ορίζουμε το σύνολο R_1 ως εξής:

$$R_1 = \left\{i | \exists b_i \in S_1 \left(d(i, b_i) = \min_{j \in S}\{d(i,j)\}\right\}\right.$$

Επομένως το συνολικό κόστος τώρα διαμορφώνεται ως εξής:

$$Cost(C_n) = a \cdot \sum_{i \in R} \min_{j \in S_1 \cup C_n}\{d(i,j)\} =$$

$$= a \cdot \sum_{i \in R_1} \min_{j \in S_1 \cup C_n}\{d(i,j)\} + a \cdot \sum_{i \in R_2} \min_{j \in S_1 \cup C_n}\{d(i,j)\} + n \cdot c_s$$

και με βάση τα παραπάνω έχουμε ότι:

$$Cost(C_n) = a \cdot \sum_{i \in R_2} \min_{j \in S_1 \cup C_n}\{d(i,j)\} + a \cdot \sum_{i \in R_1} d(i, b_i) + n \cdot c_s \quad (13)$$

Δεδομένου ότι ο δεύτερος και τρίτος όρος είναι σταθεροί, η εύρεση του n-συνδυασμού ο οποίος ελαχιστοποιεί το κόστος *Cost(C_n)* ισοδυναμεί με την εύρεση του συνδυασμού ο οποίος ελαχιστοποιεί τον πρώτο όρο, κάτι το οποίο μειώνει σημαντικά το υπολογιστικό κόστος του αλγόριθμου. Έτσι, ορίζουμε το συνολικό μεταβλητό κόστος (*VCost*) ως ακολούθως:

$$VCost(C_n) = a \cdot \sum_{i \in R_2} \min_{j \in S_1 \cup C_n}\{d(i,j)\} \quad (14)$$

και ισχυριζόμαστε ότι για κάθε $C_n \in {C_n}^{S_2}$,

$$Cost(C_n) = min\{Cost(C)\}, \Leftrightarrow VCost(C_n) = min\{VCost(C)\}.$$

Το πρόβλημά μας έχει περιοριστεί λοιπόν σε ένα πρόβλημα αναζήτησης εντός του συνόλου ${C_n}^{S_2}$. Έτσι, τα στοιχεία του συνόλου αυτού θα πρέπει να απαριθμηθούν προκειμένου να αποφασιστεί ποιο εξ αυτών ελαχιστοποιεί το συνολικό κόστος. Στη βιβλιογραφία απαντάται πληθώρα αλγόριθμων για την απαρίθμηση συνδυασμών, με μερικούς εξ αυτών να παραλληλοποιούνται [121] [122] [123] [124] [125].

7.3.2.4 Διάδοση Αντιγράφων

Το τελευταίο βήμα της διαδικασίας τοποθέτησης των αντιγράφων είναι η αντιγραφή τους από τους αποθηκευτικούς κόμβους οι οποίοι διατηρούν αντίγραφο στους αποθηκευτικούς κόμβους οι οποίοι ανήκουν στο σύνολο $\hat{C}_n$. Η διαδικασία της αντιγραφής βασίζεται σε απλή τεχνική απόδοσης βαρών στους κόμβους με βάση τη μεταξύ τους απόσταση και το τρέχον φορτίο και επιλογής αυτών με τα μεγαλύτερα βάρη.

7.3.3 Αναδιανομή Αντιγράφων

Δεδομένης της δυναμικότητας της ζήτησης των δεδομένων αλλά και της ανάγκης για διατήρηση των απαραίτητων επιπέδων ποιότητας υπηρεσίας με παράλληλη βελτιωμένη διαχείριση των αποθηκευτικών πόρων του συστήματος, είναι επιθυμητό οι αλλαγές στα σχήματα ζήτησης των δεδομένων να ανακλώνται στην τοποθεσία των αντιγράφων. Αυτό μπορεί να επιτευχθεί με την περιοδική εξέταση της ανάγκης για αναδιανομή των αντιγράφων.

Στο προτεινόμενο σύστημα, αρχικά εκτελείται ο αλγόριθμος ο οποίος παρουσιάστηκε στην παράγραφο 7.3.2 με τις ακόλουθες παραμέτρους:

- *n* ίσο με το πλήθος των υπαρχόντων αντιγράφων

- *m=0*
- *r* ίσο το πλήθος των αιτήσεων για το σύνολο δεδομένων στο ορισμένο χρονικό παράθυρο
- ο πίνακας αποστάσεων υπολογισμένος όπως και στην παράγραφο 7.3.2.

Συνακολούθως, ο αλγόριθμος θα εντοπίσει τις καλύτερες τοποθεσίες για τα υπάρχοντα αντίγραφα με βάση τις αιτήσεις δεδομένων, οι οποίες θα είναι βασισμένες σε ένα περιοδικό σχήμα, στις πιο πρόσφατες ή σε έναν ενημερωμένο μέσο όρο, αναλόγως με το επιλεγμένο σχήμα. Έστω, ότι *S* είναι το σύνολο όλων των αποθηκευτικών κόμβων του συστήματος, S_1 το σύνολο των αποθηκευτικών κόμβων οι οποίοι έχουν ήδη αντίγραφο των δεδομένων και S_2 το σύνολο των αποθηκευτικών κόμβων οι οποίοι προέκυψαν από το αποτέλεσμα του αλγόριθμου. Τότε, το σύνολο S_2 περιέχει τους αποθηκευτικούς κόμβους οι οποίοι θα διατηρούν αντίγραφο όταν η διαδικασία αναδιανομής των αντιγράφων ολοκληρωθεί.

Είναι προφανές ότι δεν είναι απαραίτητο η αναδιανομή των αντιγράφων να αφορά όλα τα αντίγραφα. Με άλλα λόγια, τα αντίγραφα τα οποία είναι αποθηκευμένα σε κόμβους οι οποίοι ανήκουν στο σύνολο $S_1 \cap S_2$ δε θα μετακινηθούν. Συνεπώς, η διαδικασία αυτή αφορά στην εύρεση του τρόπου με τον οποίο αντίγραφα από τους κόμβους οι οποίοι ανήκουν στο σύνολο $\hat{S}_1 = S_1 - S_1 \cap S_2$ θα μετακινηθούν στους κόμβους του συνόλου $\hat{S}_2 = S_2 - S_1 \cap S_2$. Αρχικά γίνεται αντιγραφή των αντιγράφων στους κόμβους του $\hat{S}_2$ με τον τρόπο ο οποίος περιγράφηκε στην παράγραφο 7.3.2.4 και εν συνεχεία διαγραφή των αντιγράφων από τους κόμβους του $\hat{S}_1$, ενώ ενημερώνεται καταλλήλως ο κατάλογος στον κόμβο διαχείρισης αντιγράφων.

7.3.4 Διαγραφή Αντιγράφων

Δεδομένου ότι η αντιγραφή δεδομένων αποτελεί μια τεχνική κατά την οποία καταναλώνονται σημαντικοί αποθηκευτικοί πόροι, τα αντίγραφα τα οποία δε χρησιμοποιούνται πια πρέπει να

διαγράφονται, προκειμένου να διατίθεται αυτός ο αποθηκευτικός χώρος πιο αποτελεσματικά, δηλαδή σε πιο «δημοφιλή» δεδομένα.

Ο αλγόριθμος ο οποίος αποφασίζει το πλήθος των δεδομένων τα οποία θα διαγραφούν αποτελεί τον δυικό αλγόριθμο του αντίστοιχου για τη δημιουργία αντιγράφων. Έτσι, ενώ ο αλγόριθμος δημιουργίας αντιγράφων ορίζει ως *S* τον χώρο ο οποίος είναι διαθέσιμος για τη δημιουργία νέων αντιγράφων, στην περίπτωση διαγραφής αυτών αποτελεί τον χώρο ο οποίος πρέπει να ελευθερωθεί.

7.3.4.1 Προσδιορισμός πλήθους αντιγράφων προς διαγραφή

Ορίζοντας τη συνάρτηση *f(i)* όπως και στην παράγραφο 7.3.1.1, η *1- f(i)* αφορά στην απουσία δημοτικότητας. Κανονικοποιώντας την τελευταία κατανομή διαιρώντας με

$$\sum_{i=1}^{k}\{1-f(i)\} = k - \sum_{i=1}^{k} f(i) = k-1$$

λαμβάνουμε τον παράγοντα:

$$g(i) = \frac{1-f(i)}{k-1} \text{ (15)}$$

Έτσι, ο χώρος $\tilde{s}(i)$ ο οποίος ελευθερώνεται με τη διαγραφή των αντιγράφων του i-οστού συνόλου δεδομένων θα πρέπει να είναι ανάλογως του *g(i)*, δηλαδή:

$$\tilde{s}(i) = g(i) \cdot S \text{ (16)}$$

και

$$s(i) = \lfloor \tilde{s}(i) \rfloor = \lfloor g(i) \cdot S \rfloor \text{ (17)}$$

Έτσι, το πλήθος των αντιγράφων τα οποία πρέπει να διαγραφούν είναι :

$$\tilde{m}(i) = s(i)/size(i) \quad (18)$$

Στο σημείο αυτό θα πρέπει να τονίσουμε ότι κατά τη διαγραφή των αντιγράφων υφίσταται ο περιορισμός ότι για κάθε δεδομένο πρέπει να υπάρχει τουλάχιστον ένα αντίγραφο. Έτσι, αν *n(i)*

είναι το πλήθος των υπαρχόντων αντιγράφων του i-οστού συνόλου δεδομένων και *m(i)* το σύνολο των αντιγράφων τα οποία θα διαγραφούν, τότε:

$$n(i) - m(i) \geq 1 \Leftrightarrow m(i) \leq n(i) - 1 \quad (19)$$

i	$size(i)$	$s(i)$	$n(i)$
i_1	$size(i_1)$	$s(i_1)$	$n(i_1)$
i_2	$size(i_2)$	$s(i_2)$	$n(i_2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
i_k	$size(i_k)$	$s(i_k)$	$n(i_k)$

Πίνακας 6 Είσοδος για τον αλγόριθμος διαγραφής αντιγράφων

Έτσι, ο αλγόριθμος για τη δημιουργία αντιγράφων προσαρμόζεται ως εξής:

I. Ταξινόμηση κατά φθίνουσα σειρά των μεγεθών στον Πίνακας 6, άρα $size(i_{\lambda+1}) \leq size(i_\lambda)$ για $\lambda = 1,\ldots,k-1$.

II. $\lambda = 1$

III. Αν $\lambda=k$, ο αλγόριθμος τερματίζει.

IV. Υπολογισμός του πλήθους των αντιγράφων τα οποία θα διαγραφούν με βάση τη σχέση:

$$m(i_\lambda) = \min(\lfloor s(i_\lambda)/size(i_\lambda) \rfloor, n(i_\lambda) - 1)$$

V. Η ποσότητα $L = s(i_\lambda) - m(i_\lambda) \cdot size(i_\lambda)$ πρέπει να μοιραστεί στα υπόλοιπα σύνολα δεδομένων. Έστω ότι $L = a(k - \lambda) + b$ είναι η ευκλείδια διαίρεση του L με το $(k - \lambda)$.

Τότε, $s(i_\mu) := s(i_\mu) + (\alpha + 1)$ για μ = λ+1,...,λ+b και $s(i_\mu) := s(i_\mu) + \alpha$ για μ = λ + b + 1,...,k

VI. λ:=λ+1

VII. Επιστροφή στο βήμα III

7.3.4.2 Επιλογή των αντιγράφων προς διαγραφή

Έστω ότι *S* είναι το σύνολο των αποθηκευτικών κόμβων οι οποίοι περιέχουν ένα αντίγραφο του συνόλου δεδομένων, |*S*| = *n*, *m* είναι το πλήθος των αντιγράφων προς διαγραφή, $C^S_{(n-m)}$ είναι το σύνολο των (*n*-*m*)-συνδυασμών από το σύνολο *S* και *R* είναι το σύνολο των αιτήσεων

προς τα δεδομένα αυτά (οι πιο πρόσφατες, οι περιοδικές ή ένας ενημερωμένος μέσος όρος). Τα *(n-m)* αντίγραφα τα οποία θα παραμείνουν θα πρέπει να είναι αυτά τα οποία ελαχιστοποιούν τη συνάρτηση κόστους και με βάση την ανάλυση στην παράγραφο 7.3.2.2 το μεταβλητό μέρος αυτής, δηλαδή:

$$VCost(C) = a \cdot \sum_{i \in R} \min_{j \in C}\{d(i,j)\} \quad (20)$$

και

$$\hat{C}_{(n-m)} = \operatorname{argmin}_{C \in C^{S}_{(n-m)}}\{VCost(C)\} \quad (21)$$

Συνεπώς, τα αντίγραφα τα οποία είναι αποθηκευμένα στους αποθηκευτικούς κόμβους του συνόλου $S - \hat{C}_{(n-m)}$ θα διαγραφούν και ο κατάλογος στον κόμβο διαχείρισης αντιγράφων θα ενημερωθεί καταλλήλως.

7.3.5 Ευριστικές

Οι αλγόριθμοι για την τοποθέτηση νέων αντιγράφων, την αναδιανομή αυτών αλλά και τη διαγραφή υπαρχόντων αντιγράφων, όπως είδαμε και στις προηγούμενες παραγράφους, έχουν αναχθεί σε ένα πρόβλημα αναζήτησης. Δεδομένου ενός συνόλου αιτήσεων *R (|R| = r),* ενός συνόλου αποθηκευτικών κόμβων *S* (*|S| = s*) από το οποίο θα γίνει η επιλογή και ενός πίνακα αποστάσεων *d(i, j),* με $i \in R$ and $j \in S$, το ζητούμενο είναι η επιλογή *n* αποθηκευτικών κόμβων τέτοιων ώστε η συνάρτηση κόστους να ελαχιστοποιείται. Όπως είδαμε και στην παράγραφο 7.3.2 στην περίπτωση τοποθέτησης νέων αντιγράφων, το j μπορεί να ανήκει σε ένα ευρύτερο σύνολο, το $S_1 \cup C_n$

Η απαρίθμηση όλων των *n*-συνδυασμών από το *S* αποτελεί μια σημαντικά απαιτητική υπολογιστικά διαδικασία για μεγάλες τιμές του *s,* ακόμη και όταν η διαδικασία είναι παραλληλοποιημένη. Τα προβλήματα αναζήτησης γενικώς είναι NP-complete. Για τον λόγο

αυτόν, θα παρουσιάσουμε και θα αξιολογήσουμε ευρυστικούς αλγόριθμους οι οποίοι βρίσκουν υποβέλτιστες λύσεις αλλά με σημαντικότατη βελτίωση του χρόνου εκτέλεσης.

7.3.5.1 Ένας καινοτόμος άπληστος αλγόριθμος με φιλτράρισμα αιτήσεων

Η βασική ιδέα της άπληστης προσέγγισης είναι η επιλογή ενός κόμβου κάθε φορά. Στην πρώτη επανάληψη, επιλέγουμε τον αποθηκευτικό κόμβο ο οποίος ελαχιστοποιεί το κόστος υπό την υπόθεση ότι όλες οι αιτήσεις εξυπηρετούνται από τον κοντινότερο κόμβο όπως φαίνεται στον πίνακα αποστάσεων. Στη δεύτερη επανάληψη, επιλέγουμε τον αποθηκευτικό κόμβο ο οποίος ελαχιστοποιεί το κόστος υποθέτοντας ότι κάθε αίτηση εξυπηρετείται από τον πιο «κοντινό» κόμβο με βάση τον πίνακα απόστασης (για κάθε αίτηση επιλέγουμε τον πιο κοντινό μεταξύ του επιλεγμένου από την πρώτη επανάληψη κόμβου και του «υποψηφίου» κόμβου από τη δεύτερη επανάληψη. Επαναλαμβάνουμε τη διαδικασία μέχρι να έχουμε επιλέξει *n* κόμβους.

Οι Qui et al [126] συγκρίνουν πολλούς αλγόριθμους τοποθέτησης αντιγράφων και καταλήγουν ότι ο άπληστος αλγόριθμος έχει πολύ καλή απόδοση. Παρόμοιες άπληστες προσεγγίσεις απαντώνται και στην έρευνα των Cronin et al [127] και έχουν αποδειχθεί αρκετά αποτελεσματικές. Για τον λόγο αυτόν χρησιμοποιούμε τον άπληστο αλγόριθμο ο οποίος παρουσιάστηκε παραπάνω ως βάση των συγκρίσεών μας και προτείνουμε μια επέκταση στον άπληστο αλγόριθμο εισάγοντας την ίδια μέθοδο φιλτραρίσματος. Στην πρώτη επανάληψη του αλγόριθμου επιλέγουμε έναν κόμβο και φιλτράρουμε τις αιτήσεις των οποίων ο «κοντινότερος» κόμβος είναι αυτός ο οποίος επελέγη. Ακολουθούμε το ίδιο φιλτράρισμα σε κάθε επανάληψη του αλγόριθμου. Είναι προφανές ότι η διαδικασία του φιλτραρίσματος συνοδεύεται από μια επιβάρυνση (overhead), η οποία όμως εξισορροπείται από τη μείωση των αιτήσεων.

Συγκεκριμένα, ορίζουμε τον «λόγο χρόνου εκτέλεσης» (“execution time ratio”) ως τον λόγο του μέσου χρόνου εκτέλεσης του άπληστου αλγόριθμου χωρίς το φιλτράρισμα προς τον μέσο

χρόνο εκτέλεσης του άπληστου αλγόριθμου με φιλτράρισμα. Έτσι, τιμές μεγαλύτερες της μονάδας δείχνουν ότι η διαδικασία φιλτραρίσματος επιταχύνει τη διαδικασία.

7.3.5.2 Ένας καινοτόμος προσαρμόσιμος ευριστικός αλγόριθμος

Ένα μειονέκτημα του άπληστου αλγόριθμου είναι ότι δε χαρακτηρίζεται από προσαρμοστικότητα, δηλαδή ανεξαρτήτως από τη διαθέσιμη υπολογιστική ισχύ, θα καταλήγει πάντοτε στο ίδιο αποτέλεσμα. Για τον λόγο αυτόν, εξετάζουμε έναν προσαρμόσιμο αλγόριθμο, η ακρίβεια των αποτελεσμάτων του οποίου ποικίλει αναλόγως με το χρονικό διάστημα κατά το οποίο επιθυμούμε να τρέξει. Με άλλα λόγια, ο αλγόριθμος αυτός θα καταλήξει στο βέλτιστο σύνολο αν δε θέσουμε χρονικό περιορισμό στην εκτέλεσή του. Έτσι, ο αλγόριθμος αυτός μπορεί να προσαρμόζεται αναλόγως με τους διαθέσιμους υπολογιστικούς πόρους.

Όπως μπορούμε να δούμε και στον Πίνακας 4, για κάθε αίτημα μπορούμε εύκολα να βρούμε τον αποθηκευτικό κόμβο b_i ο οποίος βρίσκεται πιο «κοντά». Ένα παράδειγμα παρατίθεται στον ακόλουθο πίνακα με πλήθος αιτήσεων ίσο με 10 ($r=10$) και 5 αποθηκευτικούς κόμβους ($s=5$).

	1	2	3	4	5	b_i
1	6.5	3.5	5.2	*2.3*	6.7	4
2	8.3	8.9	1.1	*0.8*	3.8	4
3	8.1	3.6	4.4	4.0	*3.3*	5
4	*5.6*	9.3	6.2	7.2	6.2	1
5	9.7	7.2	8.5	*0.2*	9.0	4
6	8.8	7.1	1.0	*0.7*	5.9	4
7	4.3	7.7	9.5	3.2	*2.6*	5
8	3.9	4.9	8.0	2.2	*1.0*	5
9	7.5	4.5	4.4	8.1	*2.3*	5
10	2.6	7.2	6.7	*2.3*	9.9	4
w_j	1	0	0	5	4	

Πίνακας 7 Πίνακας αποστάσεων για πλήθος αιτήσεων ίσο με 10 και 5 αποθηκευτικούς κόμβους

Όπως είναι λογικό με βάση την ανάλυση που έχουμε ήδη παρουσιάσει, ένας αποθηκευτικός κόμβος ο οποίος είναι «κοντά» σε ένα μεγάλο αριθμό αιτήσεων είναι πολύ πιθανό να

συμπεριληφθεί στο βέλτιστο σύνολο. Έτσι, για κάθε αποθηκευτικό κόμβο $j \in S$ μετράμε το πλήθος των αιτήσεων για τις οποίες $b_i = j$, καλώντας τον αριθμό αυτόν «βάρος» του αποθηκευτικού κόμβου και συμβολίζοντάς το με w_j. Έστω ότι στο παραπάνω παράδειγμα επιθυμούμε να επιλέξουμε τους 3 «καλύτερους» αποθηκευτικούς κόμβους. Χωρίς να κάνουμε χρήση των ευριστικών μεθόδων και απαριθμώντας τους συνδυασμούς, καταλήγουμε ότι το βέλτιστο σύνολο αποθηκευτικών κόμβων είναι το $\{1,4,5\}$, παρατηρώντας την τελευταία στήλη του Πίνακας 7.

	without heuristics		*with heuristics*	
order	combination	cost	mapped to	cost
1	1 2 3	37.2	*4 5 1*	*21.1*
2	1 2 4	25.4	4 5 3	21.7
3	1 2 5	37.8	4 5 2	21.7
4	1 3 4	25.7	4 1 3	25.7
5	1 3 5	33.2	4 1 2	25.4
6	*1 4 5*	*21.1*	4 3 2	25.9
7	2 3 4	25.9	5 1 3	33.2
8	2 3 5	34.9	5 1 2	37.8
9	2 4 5	21.7	5 3 2	34.9
10	3 4 5	21.7	1 3 2	37.2

Πίνακας 8 Εύρεση του βέλτιστου συνόλου αποθηκευτικών κόμβων με και δίχως ευριστικές

storage node	1	2	3	4	5
weight	1	0	0	5	4
weight order	3	5	4	1	2
weight order	1	2	3	4	5
weight	5	4	1	0	0
storage node	4	5	1	3	2

Πίνακας 9 Αποθηκευτικοί κόμβοι ταξινομημένοι κατά φθίνουσα σειρά βάρους

Στον Πίνακας 8 παρατηρούμε ότι το βέλτιστο σύνολο είναι ο $6^{ος}$ συνδυασμός στην ταξινομημένη λεξικογραφικά λίστα. Δεν είναι τυχαίο το γεγονός ότι οι αποθηκευτικοί κόμβοι με το μεγαλύτερο βάρος w_j είναι αυτοί οι οποίοι ανήκουν στο βέλτιστο σύνολο.

Η ιδεά της προσέγγισής μας είναι η ταξινόμηση των αποθηκευτικών κόμβων κατά φθίνουσα σειρά βάρους (w_j) έτσι ώστε οι αποθηκευτικοί κόμβοι με το μεγαλύτερο βάρος να είναι οι πρώτοι στη λίστα των συνδυασμών (βλ. Πίνακας 9). Χρησιμοποιώντας την αντιστοίχηση η οποία ορίζεται στον πίνακα αυτόν, οι συνδυασμοί απαριθμούνται στην επιθυμητή σειρά, όπως μπορούμε να δούμε και στις δύο τελευταίες στήλες του Πίνακας 8. Το βέλτιστο σύνολο αποθηκευτικών κόμβων {1,4,5} τώρα εμφανίζεται στην πρώτη θέση της λίστας των συνδυασμών. Προφανώς, η προσαρμοστικότητα του αλγόριθμου έγκειται στο γεγονός ότι μπορούμε να περιορίσουμε το πλήθος των συνδυασμών τους οποίους ελέγχει. Έτσι μπορούμε να ελέγξουμε μόνο τον πρώτο συνδυασμό (και στην περίπτωση αυτή ο προτεινόμενος αλγόριθμος φέρει ομοιότητες με τις Hot spot ευριστικές [126] ή τους πρώτους s ή s^2 συνδυασμούς, μεταξύ των άλλων.

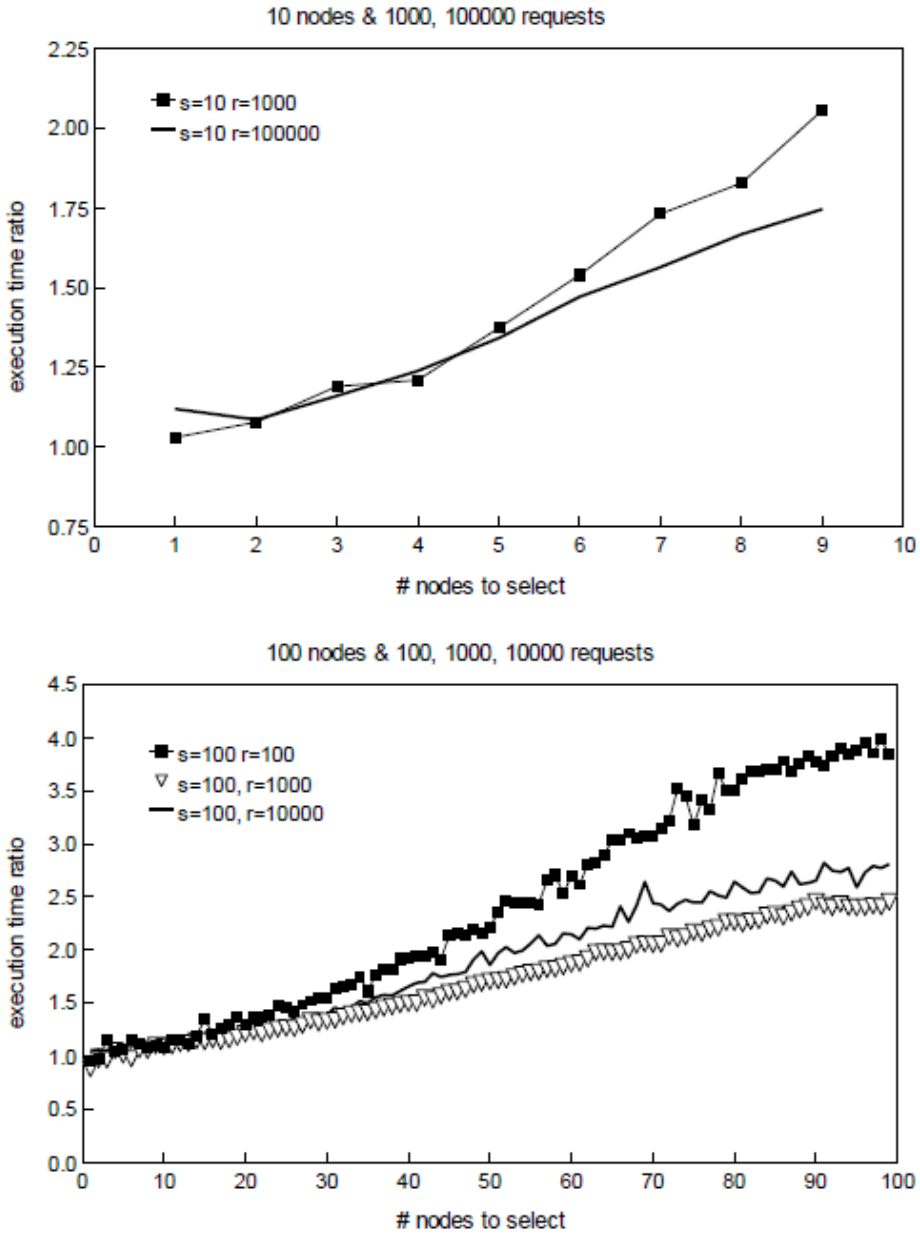
7.4 *Αξιολόγηση προτεινόμενων αλγόριθμων*

Στην παράγραφο αυτή θα αξιολογήσουμε τους παρουσιασθέντες ευριστικούς αλγόριθμους.

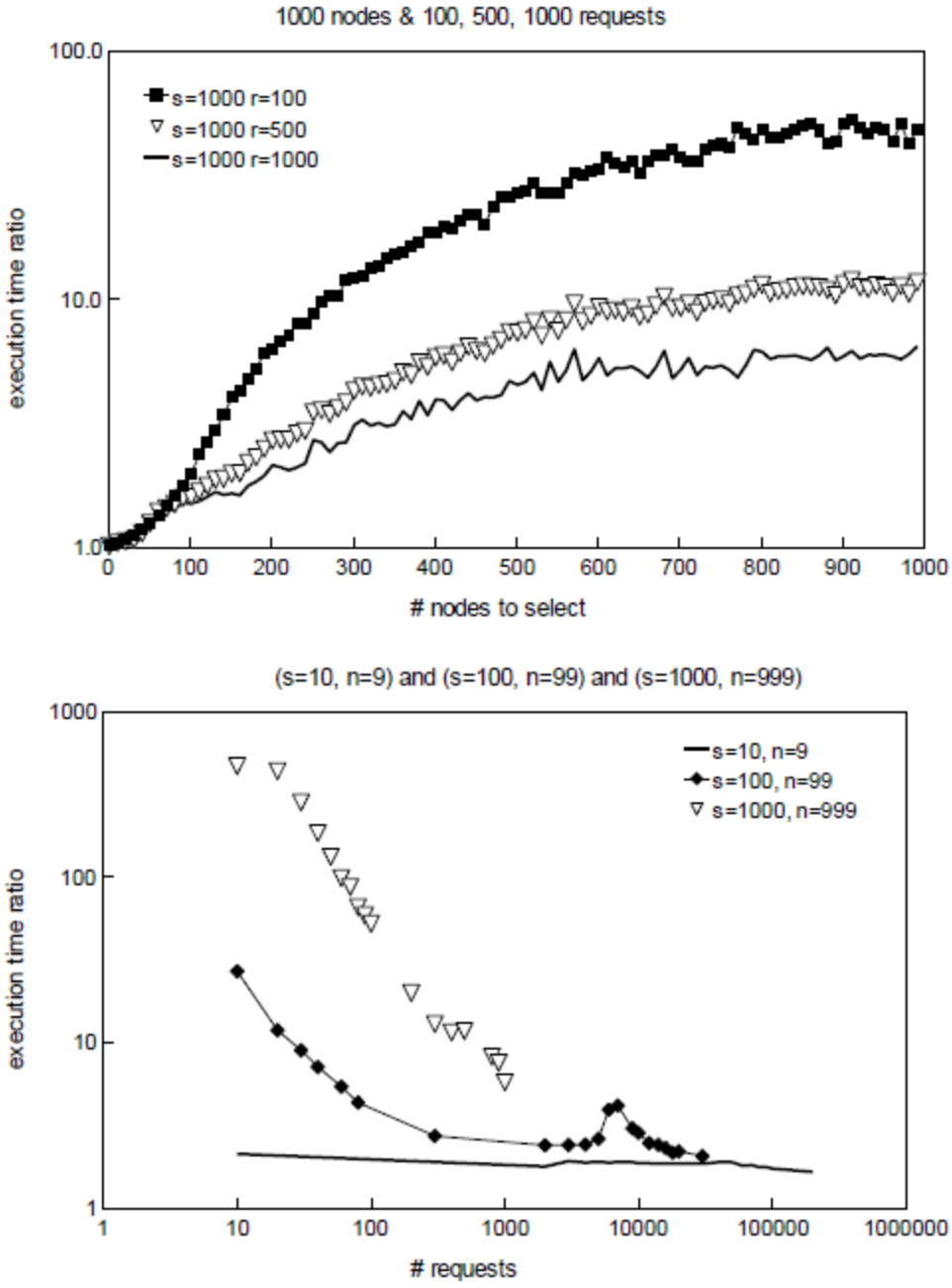
7.4.1 Άπληστος Ευριστικός Αλγόριθμος

Για την αξιολόγηση του ευριστικού αλγόριθμου με φιλτράρισμα των αιτήσεων εστιάσαμε την αξιολόγησή μας σε ορισμένα πεδία τιμών των *s, n* και *r* τα οποία φαίνονται στην Εικόνα 17 και στην Εικόνα 18. Έτσι, αρχικά θεωρήσαμε 10 αποθηκευτικούς κόμβους και υπολογίσαμε τον λόγο χρόνου εκτέλεσης συναρτήσει του πλήθους των αντιγράφων προς δημιουργία *n* για διαφορετικές τιμές πλήθους αιτήσεων. Ομοίως, ακολούθησαν αντίστοιχοι υπολογισμοί για 100 και 1000 αποθηκευτικούς κόμβους. Επιπροσθέτως, έλαβε χώρα ο υπολογισμός του λόγου χρόνου εκτέλεσης συναρτήσει του πλήθους των αιτήσεων για διαφορετικές τιμές του ζεύγους (*s*,*n*). Για κάθε σημείο στα διαγράμματα αυτά έλαβαν χώρα 500 πειράματα προκειμένου να

προκύψουν αξιόπιστες μέσες τιμές των μεγεθών. Όπως μπορούμε να παρατηρήσουμε και στις ακόλουθες γραφικές παραστάσεις ο αλγόριθμος με φιλτράρισμα αιτήσεων είναι σε όλες τις περιπτώσεις σημαντικά πιο γρήγορος (περισσότερο από δύο φορές πιο γρήγορος) από τον απλό άπληστο αλγόριθμο.



Εικόνα 17 Σύγκριση ευριστικού αλγόριθμου με και χωρίς φιλτράρισμα (s=100)



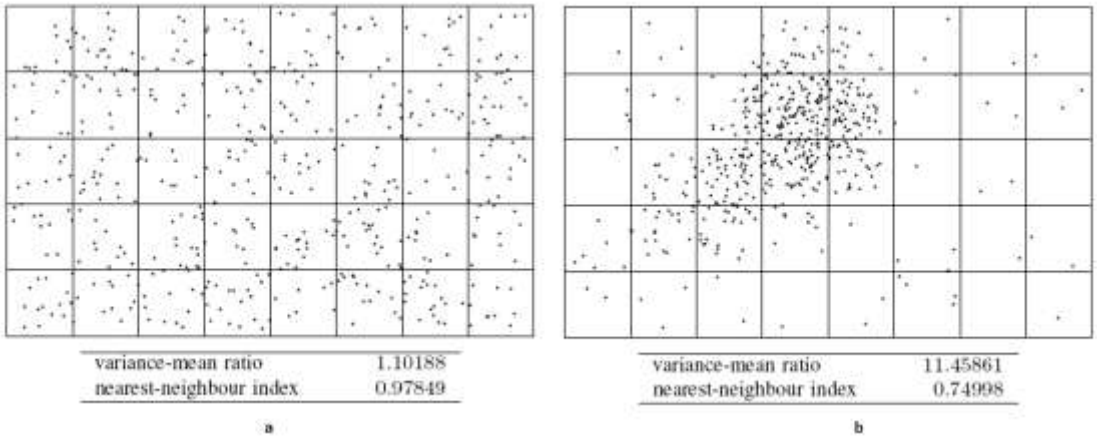
Εικόνα 18 Σύγκριση ευριστικού αλγόριθμου με και χωρίς φιλτράρισμα για α) s=1000 και β) διάφορες τιμές του *s* και του *n*

7.4.2 Η επίδραση των σχημάτων των αιτήσεων στον ευριστικό αλγόριθμο

Έστω ότι το σχήμα των αιτήσεων είναι αρκετά ξεκάθαρο, δηλαδή οι αιτήσεις «κοντά» σε συγκεκριμένους αποθηκευτικούς κόμβους. Προφανώς τότε, τα βάρη των αποθηκευτικών

κόμβων αυτών θα είναι αρκετά υψηλά. Έτσι, ο προτεινόμενος ευριστικός αλγόριθμος θα καταλήξει στο βέλτιστο σύνολο σχετικά σύντομα. Θα αποδείξουμε πειραματικά τους ισχυρισμούς αυτούς.

Θα χρησιμοποιήσουμε quadrat και nearest-neighbour ανάλυση σχήματος σημείων για να χαρακτηρίσουμε τα σχήματα των αιτήσεων. Η quadrat ανάλυση καταλήγει σε μια μετρική η οποία ονομάζεται λόγος διακύμανσης-μέσης τιμής (variance-mean ratio), η οποία αποτελεί ένα μέτρο της διασποράς των σημείων. Τιμές κοντά στη μονάδα δείχνουν τυχαία διεσπαρμένα σημεία, ενώ τιμές μεγαλύτερες τις μονάδας δείχνουν ομαδοποίηση. Ομοίως η nearest-neighbour ανάλυση καταλήγει σε μια μετρική η οποία καλείται δείκτης πλησιέστερου γείτονα, η οποία αφορά στον λόγο της μέσης απόστασης των σημείων από τον πιο κοντινό τους γείτονα προς τη μέση απόσταση η οποία θα αναμενότας από μια τυχαία κατανομή τους. Έτσι, τιμές κοντά στη μονάδα δείχνουν τυχαία σχήματα, ενώ τιμές μικρότερες τις μονάδας δείχνουν σχήματα ομαδοποίησης.



Εικόνα 19 α) Τυχαίο σχήμα σημείων σε δισδιάστατο χώρο β) Σχήμα ομαδοποίησης σημείων σε δισδιάστατο χώρο

Στην Εικόνα 19α) μπορούμε να δούμε ένα σχεδόν πλήρως τυχαίο σχήμα αιτήσεων και στην Εικόνα 19β) παρατηρούμε ένα σχήμα ομαδοποίησης αιτήσεων. Θα επιλέξουμε ως τυχαία

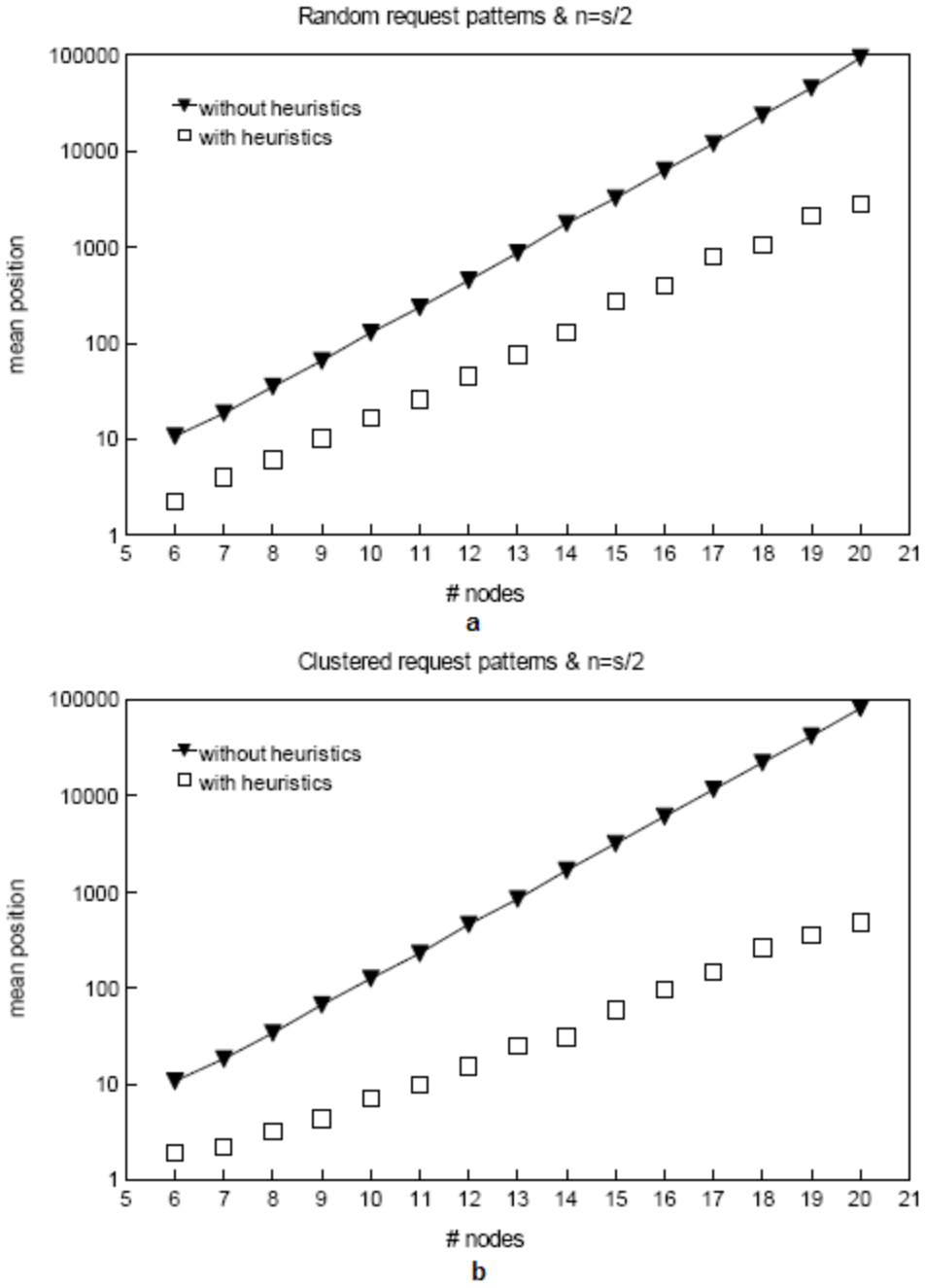
σχήματα αυτά τα οποία έχουν δείκτη πλησιέστερου γείτονα (Nearest0Neighbour Index – NNI) στο διάστημα (0.98-1.02) και ως σχήματα ομαδοποίησης αυτά με δείκτη πλησιέστερου γείτονα στο διάστημα (0.70-0.75).

Το κριτήριο για την αξιολόγηση του ευριστικού αλγόριθμου θα είναι *η θέση του πρώτου βέλτιστου συνόλου στη λίστα συνδυασμών*. Στο παράδειγμα που είδαμε στην παράγραφο 7.3.5.2 είδαμε ότι χωρίς τη χρήση ευριστικών το βέλτιστο σύνολο είναι στην $6^{η}$ θέση της λίστας συνδυασμών, ενώ με τη χρήση ευριστικών βρέθηκε στην πρώτη θέση.

Αρχικά θα εξετάσουμε τον τρόπο με τον οποίο η θέση του πρώτου βέλτιστου συνόλου επηρεάζεται από το πλήθος των διαθέσιμων αποθηκευτικών κόμβων και το πλήθος των αποθηκευτικών κόμβων οι οποίοι έχουν επιλεγεί για την τοποθέτηση των νέων αντιγράφων (s και n αντιστοίχως). Στο σημείο αυτό πρέπει να παρατηρήσουμε ότι η χειρότερη περίπτωση είναι όταν το n είναι στη μέση των επιτρεπόμενων τιμών του (οι οποίες είναι από 0 ως s). Πράγματι,

$$\max_{n=0,\dots,s} \binom{s}{n} = \binom{s}{\lfloor s/2 \rfloor}$$

Εξετάζοντας τη χειρότερη περίπτωση, δηλαδή $n = \lfloor s/2 \rfloor$ χρειαζόμαστε μόνο ένα δισδιάστατο διάγραμμα. Στο παράδειγμά μας $s \in \{6, \dots, 20\}$.



Εικόνα 20 Μέση θέση του πρώτου βέλτιστου συνόλου με και δίχως τη χρήση ευριστικών για τυχαία σχήματα και σχήματα ομαδοποίησης

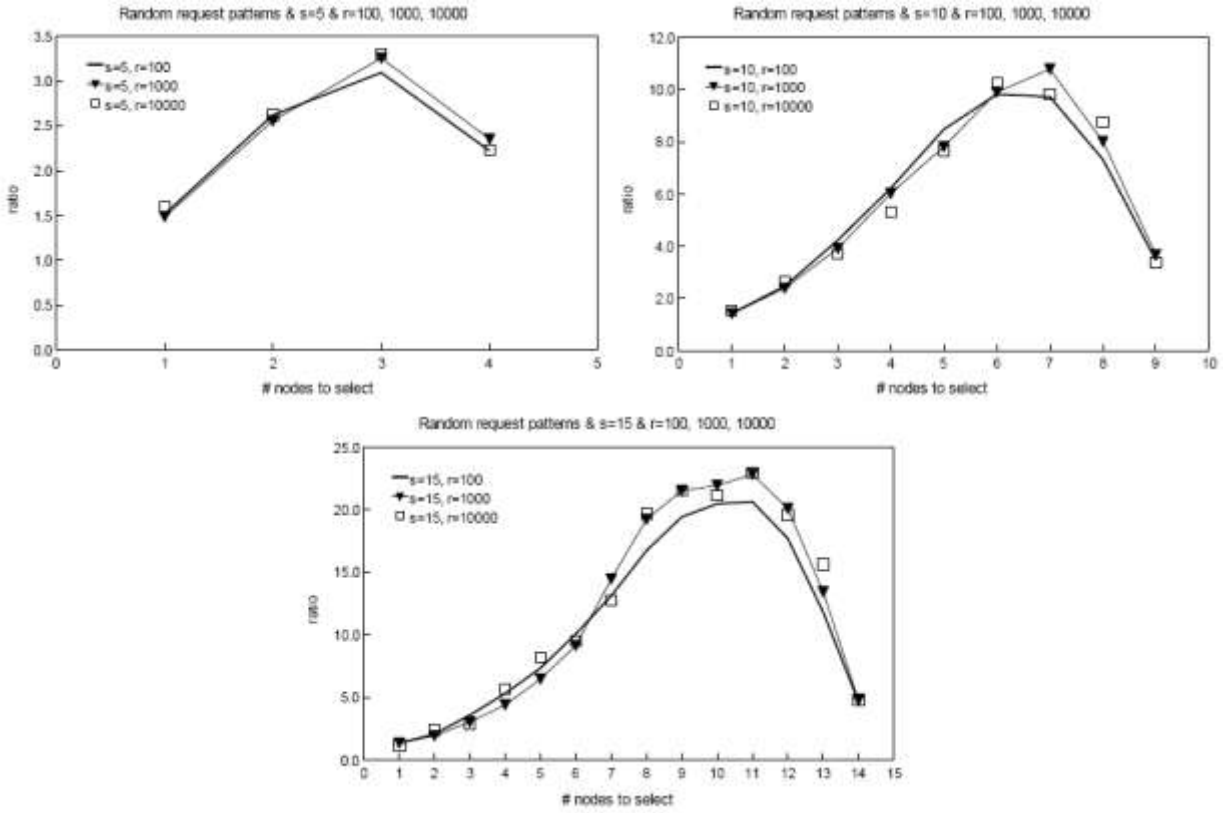
Για κάθε σημείο στα διαγράμματα της Εικόνα 20 έλαβαν χώρα 500 πειράματα. Έστω ότι η μέση θέση του βέλτιστου συνόλου στη λίστα συνδυασμών χωρίς τη χρήση ευριστικών να είναι

$\overline{p}^{\,without}$ και με τη χρήση ευριστικών $\overline{p}^{\,with}$. Προκειμένου να συγκρίνουμε την απόδοση του ευριστικού αλγόριθμου με τον μη ευριστικό αλγόριθμο ορίζουμε τον λόγο

$$\overline{p}^{\,without}/\overline{p}^{\,with}$$

Μεγαλύτερες τιμές του λόγου σημαίνουν καλύτερη απόδοση της ευριστικής προσέγγισης. Το πρώτο διάγραμμα στην Εικόνα 20 δείχνει ότι για τυχαία σχήματα αιτήσεων η προτεινόμενη ευριστική προσέγγιση προσφέρει σημαντική βελτίωση στη θέση του πρώτου βέλτιστου συνόλου. Επιπροσθέτως, παρατηρούμε ότι για αυξανόμενες τιμές του s η απόσταση των δύο καμπυλών αυξάνει και συνεπώς ο λόγος $\overline{p}^{\,with}/\overline{p}^{\,without}$ ακολουθεί αύξουσα πορεία. Στο δεύτερο διάγραμμα της Εικόνα 20 παρατηρούμε ότι για σχήματα ομαδοποίησης ο προτεινόμενος ευριστικός αλγόριθμος αποδίδει ακόμη καλύτερα.

Θα πρέπει να τονίσουμε ότι, όπως μπορούμε να δούμε και στην Εικόνα 21 ο λόγος $\overline{p}^{\,with}/\overline{p}^{\,without}$ είναι ανεξάρτητος από το πλήθος των αιτήσεων. Για κάθε σημείο στα διαγράμματα της Εικόνα 21 έλαβαν χώρα 1000 πειράματα.



Εικόνα 21 Σχέση μεταξύ λόγου $\overline{p}^{without}/\overline{p}^{with}$ και πλήθους αποθηκευτικών κόμβων

7.4.3 Σύγκριση προσαρμοστικού ευριστικού αλγόριθμο με τον άπληστο ευριστικό αλγόριθμο

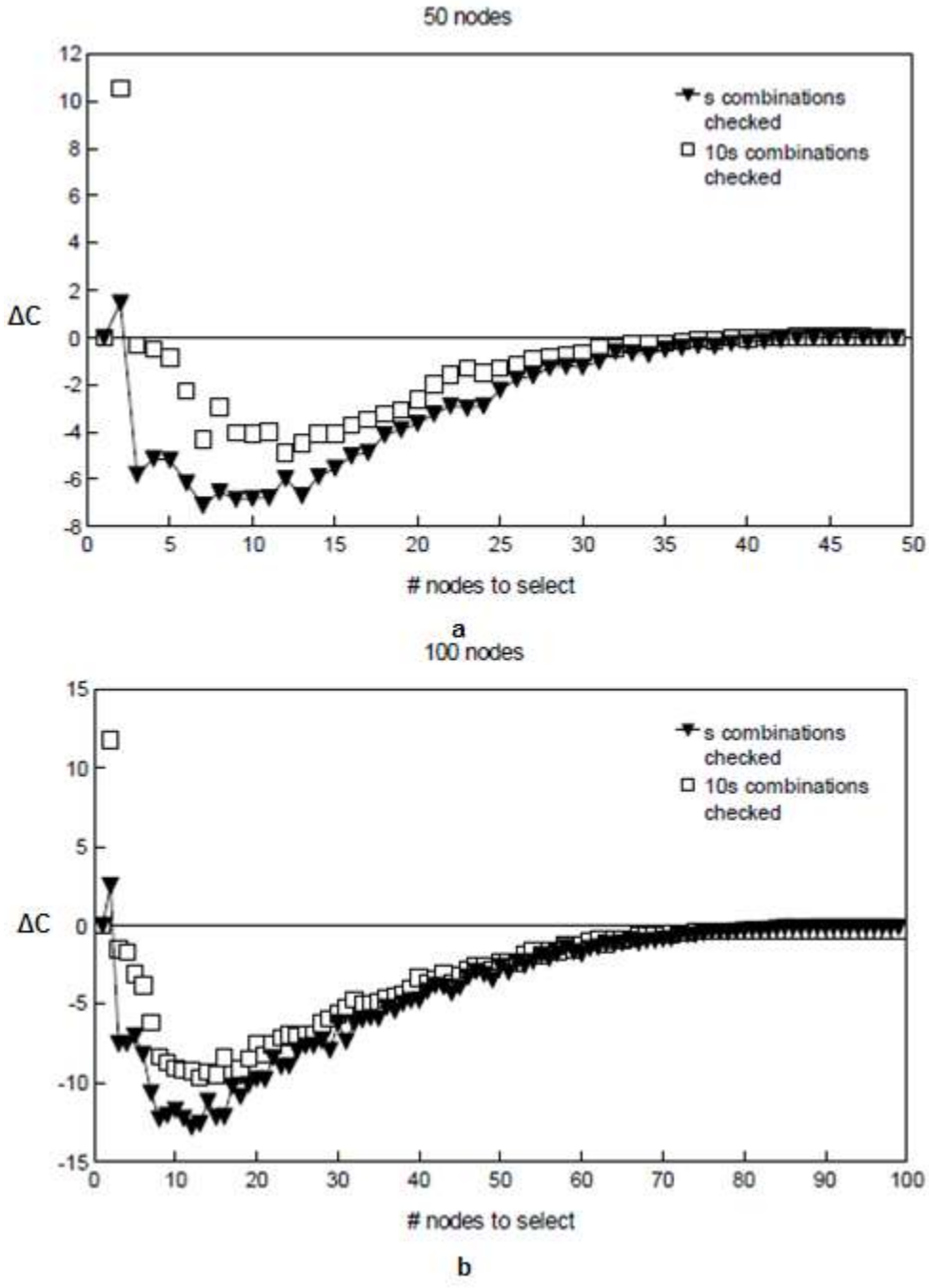
Κάθε ευριστικός αλγόριθμος καταλήγει σε έναν συνδυασμό ο οποίος συνοδεύεται από ένα συγκεκριμένο κόστος. Δεδομένων δύο συνδυασμών, θεωρούμε καλύτερο αυτόν με το μικρότερο κόστος. Έστω $cost_1$ το κόστος του συνδυασμού ο οποίος προκύπτει από τον άπληστο αλγόριθμο και $cost_2$ το κόστος του συνδυασμού ο οποίος προκύπτει από τον προσαρμόσιμο αλγόριθμο ο οποίος παρουσιάστηκε στην παράγραφο 7.3.5.2. Προκειμένου να συγκρίνουμε τους δύο αυτούς αλγόριθμους θα εισαγάγουμε την ποσότητα ΔC:

$$\Delta C = \frac{cost_1 - cost_2}{cost_2} \cdot 100\%$$

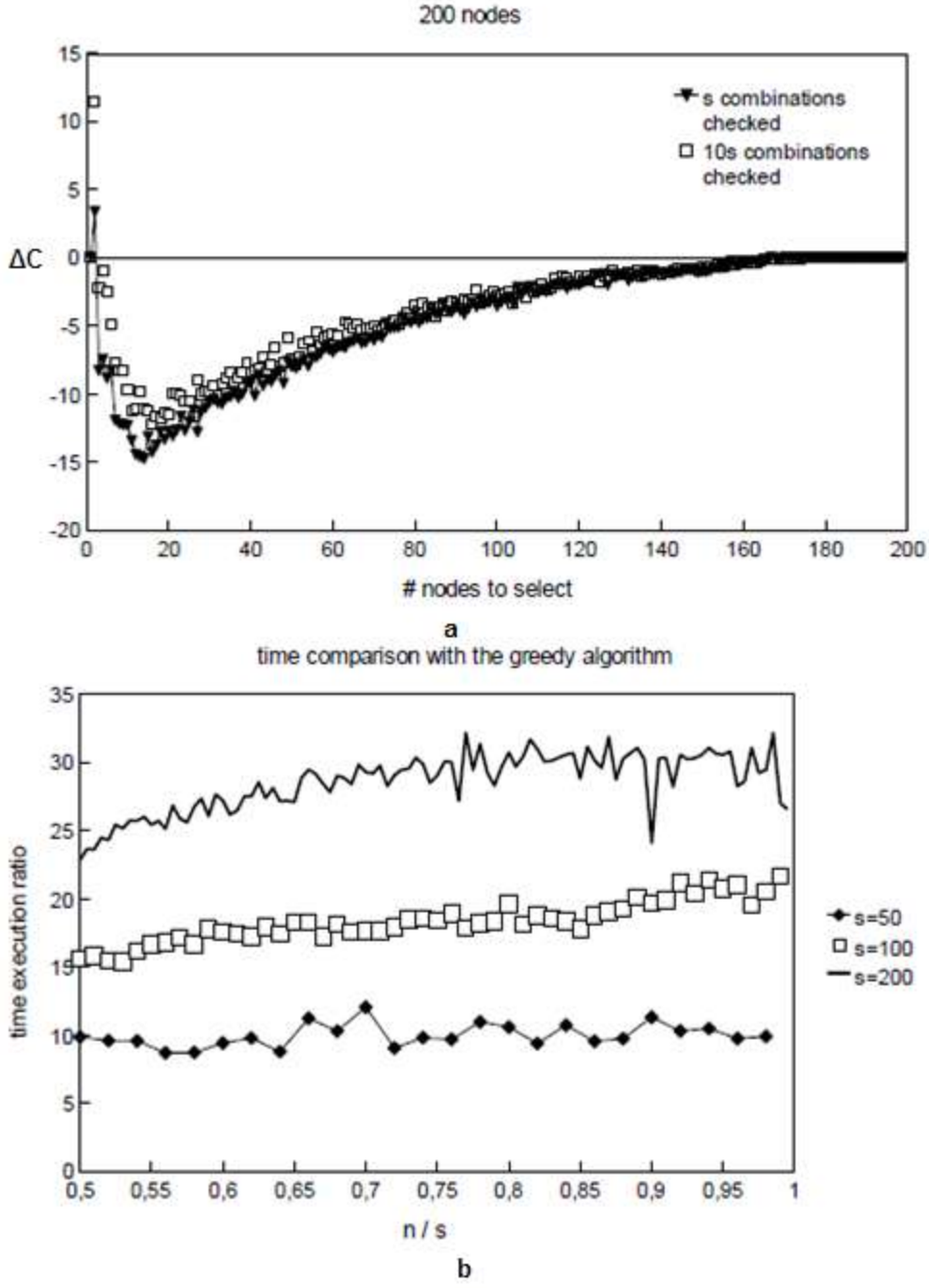
η οποία δείχνει το ποσοστό κατά το οποίο είναι μεγαλύτερο το κόστος του συνδυασμού ο οποίος είναι το αποτέλεσμα του άπληστου αλγόριθμου από το αντίστοιχο του προσαρμόσιμου ευριστικού. Έτσι αν $\Delta C>0$, τότε ο άπληστος αλγόριθμος είναι λιγότερο αποτελεσματικός από τον προτεινόμενο προσαρμοστικό ευριστικό και αντιστρόφως. Τιμές κοντά στο *0* δείχνουν ότι οι δύο αλγόριθμοι έχουν σχεδόν την ίδια αποτελεσματικότητα. Στα διαγράμματα a και b της Εικόνα 22 και στο διάγραμμα c της Εικόνα 23 παρουσιάζεται η μεταβολή του ΔC συναρτήσει του πλήθους των αντιγράφων προς δημιουργία *n* για τον προσαρμοστικό ευριστικό αλγόριθμο και τον άπληστο ευριστικό αλγόριθμο για διαφορετικές τιμές του πλήθους των αποθηκευτικών κόμβων *s* για ελεγμένους *s* και *10s* συνδυασμούς από τον προσαρμοστικό αλγόριθμο. Για κάθε σημείο στα διαγράμματα αυτά έλαβαν χώρα 1000 πειράματα. Παρατηρούμε ότι ο άπληστος αλγόριθμος δίνει καλύτερα αποτελέσματα για μικρές σχετικά τιμές του πλήθους των νέων αντιγράφων προς δημιουργία *n*, ενώ οι δύο αλγόριθμοι συγκλίνουν για διαρκώς αυξανόμενες τιμές του *n*. Μάλιστα, συγκρίνοντας τα αποτελέσματα αυτά παρατηρούμε ότι για μικρές τιμές του λόγου *n/s* ο άπληστος αλγόριθμος αποδίδει καλύτερα, ενώ για μεγάλες τιμές του λόγου αυτού αποδίδουν εξ ίσου αποτελεσματικά. Στο σημείο αυτό αξίζει να παρατηρήσουμε ότι για n/s > 0.5 (δηλαδή για n>s/2) το ΔC είναι πολύ κοντά στο μηδέν.

Ο προτεινόμενος προσαρμοστικός αλγόριθμος, όμως, είναι πολύ πιο αποδοτικός από τον άπληστο ευριστικό αλγόριθμο όσον αφορά στον απαιτούμενο υπολογιστικό χρόνο. Η σύγκριση γίνεται με τον λόγο χρόνου εκτέλεσης, όπως ορίστηκε ανωτέρω. Όπως μπορούμε να παρατηρήσουμε στην Εικόνα 23, ειδικά για τιμές του λόγου *n/s* μεγαλύτερες του *0.5* ο προτεινόμενος αλγόριθμος είναι *10* φορές πιο γρήγορος από τον άπληστο αλγόριθμο για *s=50*, περίπου 15-20 φορές για *s=100* και περίπου *s=25-30* φορές για *s=200*. Το γεγονός ότι ο λόγος χρόνου εκτέλεσης αυξάνεται όσο αυξάνει το πλήθος των αποθηκευτικών κόμβων (*s*), δεδομένου

ότι ο απαιτούμενος υπολογιστικός χρόνος για τον άπληστο αλγόριθμο αυξάνει δραματικά για μεγάλες τιμές του s και του n.



Εικόνα 22 Σύγκριση προσαρμοστικού ευριστικού με άπληστο ευριστικό για a) 50 και b) 100 αποθηκευτικούς κόμβους



Εικόνα 23 Σύγκριση προσαρμοστικού ευριστικού με άπληστο ευριστικό a) για 200 αποθηκευτικούς κόμβους και b) ως προς τον χρόνο εκτέλεσής τους για *s=50,100 και 200*

7.5 *Συμπεράσματα*

Στον παρόν κεφάλαιο παρουσιάσαμε ένα σύνολο καινοτόμων παραλληλίσιμων αλγόριθμων οι οποίοι προσφέρουν μια ολοκληρωμένη λύση για τη δυναμική διαχείριση αντιγράφων βιομετρικών δεδομένων σε ένα κατανεμημένο περιβάλλον. Το σύστημα αυτό ενσωματώνει ένα πλήθος λειτουργιών, όπως ο καθορισμός της χρονικής στιγμής κατά την οποία ένα νέο αντίγραφο πρέπει να δημιουργηθεί, του πλήθους των αντιγράφων τα οποία απαιτούνται με βάση το ακολουθούμενο σχήμα ζήτησης των δεδομένων, της τοποθεσίας των νέων αντιγράφων στο κατανεμημένο περιβάλλον βάση της γεωγραφίας των αιτήσεων για δεδομένα και του συνεπαγόμενου κόστους καθώς και της αναδιανομής των αντιγράφων και της διαγραφής των αντιγράφων με βάση το ενημερωμένο σχήμα ζήτησης των δεδομένων. Η προτεινόμενη λύση δυναμικής διαχείρισης αντιγράφων μπορεί να χειριστεί τη δυναμικότητα του κατανεμημένου βιομετρικού περιβάλλοντος με την επέκταση ή τη μείωση του συνόλου των αντιγράφων των βιομετρικών δεδομένων με βάση το σχήμα της ζήτησης αυτών, του κόστους αποθήκευσης, πρόσβασης και διατήρησης και της γεωγραφίας των αιτήσεων.

8 *Σύνοψη Κυριότερων Συνεισφορών - Ανοιχτά Ερευνητικά Θέματα*

Στο συγκεκριμένο κεφάλαιο περιλαμβάνεται η σύνοψη της διατριβής και τα συμπεράσματα που εξήχθησαν κατά την εκπόνησή της καθώς και η κύρια συνεισφορά της στο χώρο της τεχνολογίας πλέγματος και συζητούνται θέματα μελλοντικής εργασίας και επέκτασης των ερευνητικών αποτελεσμάτων.

8.1 *Σύνοψη Διατριβής και Κυριότερων Συνεισφορών*

Η παρούσα διατριβή αφορούσε στην παρουσίαση καινοτόμων μηχανισμών οι οποίοι στοχεύουν στη βελτίωση της αποτελεσματικότητας και της απόδοσης των βιομετρικών συστημάτων. Όσον αφορά στην αποτελεσματικότητα των τελευταίων, η έρευνα εστίασε στην ανάπτυξη μιας μοντελοποίησης των δεδομένων εκπαίδευσης των βιομετρικών συστημάτων σε επίπεδο βαθμολογίας σύγκρισης βασισμένης σε Student t mixture models και στην εφαρμογή

ενός πιθανοτικού κανόνα απόφασης ο οποίος εφαρμόζεται κατά τη φάση επερώτησης των βιομετρικών συστημάτων για τον καθορισμό της εγκυρότητας της ταυτότητας των χρηστών. Ο πειραματικός έλεγχος των μοντέλων αυτών βασίστηκε σε δεδομένα της βάσης NIST-BSSR1 [110] η οποία παρείχε δεδομένα βαθμολογίας από δύο διαφορετικά συστήματα αναγνώρισης προσώπου και από ένα σύστημα αναγνώρισης δαχτυλικών αποτυπωμάτων στο οποίο τα βιομετρικά δεδομένα ήταν δύο ειδών: δαχτυλικά αποτυπώματα από τον δεξιό δείκτη και από τον αριστερό δείκτη του χεριού. Σύμφωνα με τα αποτελέσματα των δοκιμών, η προτεινόμενη μέθοδος παρέχει σημαντική βελτίωση τόσο συγκριτικά με την απόδοση των επιμέρους συστημάτων όσο και σε σχέση με υπάρχουσες μεθοδολογίες, όπως είναι τα GMM.

Στα πλαίσια της έρευνας για την αύξηση της ακρίβειας των αποτελεσμάτων των βιομετρικών δεδομένων, μελετήθηκε και παρουσιάστηκε μια καινοτόμος αρχιτεκτονική των βιομετρικών συστημάτων η οποία αποτελεί επέκταση της υπάρχουσας γενικής αρχιτεκτονικής και η οποία στοχεύει στην πιο αποτελεσματική και έξυπνη εκμετάλλευση του πληροφοριακού πλούτου των βιομετρικών δειγμάτων τα οποία υποβάλλονται προς επερώτηση σε ένα βιομετρικό σύστημα μέσω του συνδυασμού αυτής με αντίστοιχη πληροφορία η οποία αφορά στις ειδικές συνθήκες βέλτιστης απόδοσης των εισηγμένων αλγόριθμων εξαγωγής βιομετρικών χαρακτηριστικών στο βιομετρικό σύστημα. Προχωρόντας ένα βήμα πιο πέρα, η μελέτη αυτή επεκτάθηκε σε θέματα ασφάλειας και ιδιωτικότητας των βιομετρικών συστημάτων κατά την οποία έλαβε χώρα ανάλυση των τρωτών σημείων των βιομετρικών συστημάτων και αντιστοίχηση αυτών με τις κύριες εκφάνσεις της ιδιωτικότητας.

Τέλος, βασιζόμενοι στο γεγονός ότι η ίδια η πληροφορία αποτελεί τον πιο σημαντικό πόρο σε ένα βιομετρικό σύστημα, με τα ζητήματα αξιοπιστίας, αποτελεσματικότητας και αποδοτικότητας του τελευταίου να είναι άρρηκτα συνδεδεμένα με τη διαθεσιμότητα των βιομετρικών δεδομένων

και την ταχύτητα πρόσβασης σε αυτά, αναπτύχθηκαν καινοτόμοι μηχανισμοί δημιουργίας και διατήρησης αντιγράφων σε κατανεμημένο περιβάλλον καθώς και η αξιολόγηση αυτών, με κυριότερη έμφαση στην απαιτούμενη υπολογιστική ισχύ αυτών και την ταχύτητα εκτέλεσής τους χωρίς να μειώνεται η αποτελεσματικότητά τους σε μη αποδεκτά επίπεδα.

8.2 *Ανοικτά ερευνητικά θέματα και Μελλοντική Έρευνα*

Πέραν των θεμάτων απόδοσης των βιομετρικών συστημάτων όπως είναι η αποτυχία να αποκτηθούν βιομετρικά δεδομένα καθώς και να εξαχθούν εξ αυτών βιομετρικά χαρακτηριστικά και τα σφάλματα ταυτοποίησης και επιβεβαίωσης ταυτότητας τα οποία είδαμε στην παράγραφο 4.3, οι βιομετρικές τεχνολογίες αντιμετωπίζουν πληθώρα προκλήσεων σε επίπεδο λειτουργίας, απόδοσης και αποτελεσματικότητας, οι οποίες και αναλύονται ακολούθως.

Δεδομένου ότι τα βιομετρικά συστήματα συγκεντρώνουν πλήθος τεχνολογιών, από μεθόδους απόκτησης βιομετρικών δεδομένων όπως ανάγνωση δακτυλικών αποτυπωμάτων και μαθηματικές και στατιστικές μεθόδους επεξεργασίας αυτών, μέχρι διασύνδεση των τεχνολογιών αυτών, ασφάλεια και ιδιωτικότητα. Με άλλα λόγια συνιστούν ένα ερευνητικό πεδίο το οποίο απαιτεί την ενεργό και αγαστή *συνεργασία διαφορετικών πεδίων*, όπως είναι η επιστήμη υπολογιστών, οι τεχνολογίες αισθητήρων, η ιατρική, οι κοινωνικές επιστήμες και η νομική πλευρά.

Μια σημαντική απαίτηση ενός βιομετρικού συστήματος μεγάλης κλίμακας είναι και η *ανταλλαγή δεδομένων*. Όπως έχουμε ήδη αναφέρει, τα περισσότερα βιομετρικά συστήματα τα οποία βρίσκονται στην αγορά είναι περιορισμένης κλίμακας και αφορούν έναν συγκεκριμένο οργανισμό. Επιπροσθέτως, υπάρχουν βιομετρικά συστήματα τα οποία έχουν διάφορες αρχιτεκτονικές για διαφορετικά σενάρια, ενώ τα ανεξάρτητα βιομετρικά συστήματα είναι

περιορισμένου σκοπού κυρίως λόγω της ανεπαρκούς τους ακρίβειας ή της μη αποδοχής τους από τους χρήστες σε ένα εύρος σεναρίων. Υπάρχει λοιπόν ένα σημαντικό ζητούμενο σχετικά με την *αποτελεσματικότητα* όλων αυτών των συστημάτων: η εύρεση ενός τρόπου συνεργασίας τους η οποία θα επιτρέπει την *επικοινωνία μεταξύ ετερογενών συστημάτων* και παράλληλα θα λειτουργεί υπό τις απαιτούμενες συνθήκες *ασφάλειας* και *ιδιωτικότητας*.

Το κόστος της εγκατάστασης βιομετρικών συστημάτων μειώνεται. Όμως, η ανάπτυξη μιας αποτελεσματικής και αποδοτικής βιομετρικής λύσης θέτει υψηλές απαιτήσεις στις υποδομές για επεκτασιμότητα, αξιοπιστία, απόδοση, ασφάλεια ειδικά στην περίπτωση συστημάτων μεγάλης κλίμακας και/ή πραγματικού χρόνου οδηγώντας σε σημαντικός κόστος και πολύπλοκες ή αδύναμες λύσεις. Δεδομένης της ευαισθησίας των δεδομένων και του επιπέδου κινδύνου από την έκθεση αυτών, απαιτούνται προηγμένες τεχνικές διαχείρισης συμφωνιών διασφάλισης επιπέδου υπηρεσιών (Service Level Agreements – SLAs) και διαχείρισης δεδομένων.

Λαμβάνοντας υπ’οψιν ότι ένα βιομετρικό σύστημα μεσαίας ή ευρείας κλίμακας έχει υψηλές υπολογιστικές απαιτήσεις – ειδικά στην περίπτωση λειτουργίας αυτού σε πραγματικό χρόνο και στην περίπτωση ταυτοποίησης – αλλά και λαμβάνοντας υπ’όψιν τόσο τους χρόνους «αργίας» των πόρων του συστήματος (οι οποίοι είναι άμεσα εξαρτώμενοι από την εφαρμογή του συστήματος) όσο και το συνοδευόμενο κόστος (τόσο του εξοπλισμού όσο και της συνεχούς δέσμευσης των πόρων αυτού), προσφάτως ξεκίνησαν προσπάθειες για την ανάπτυξη του *Βιομετρικού Πλέγματος* (*Biometrics Grid*) [102], το οποίο αποτελεί ένα εξειδικευμένο σύστημα βασισμένο σε μεσισμικό πλέγματος (grid middleware) για βιομετρικές εφαρμογές. Κύριος στόχος του Βιομετρικού Πλέγματος είναι να κάνει δυνατή τη συνεργασία συστημάτων, να παρέχει δυνατότητες ανταλλαγής δεδομένων και να αυξήσει την αποδοτικότητα των συστημάτων αυτών.

Στην εργασία τους οι Ming και Ma [102] προτείνουν τη χρήση των τεχνολογιών πλέγματος προκειμένου να κατανεμηθούν τα υπολογιστικά βήματα ενώ παράλληλα ικανοποιούνται οι απαιτήσεις για Ποιότητα Υπηρεσίας. Έτσι, στο σύστημα τους, το BMG (Biometrics Grid), τοποθετούν βιομετρικά Web services. Θεωρούν ένα σύνολο από M εφαρμογές $A = \{a_1, a_2, \ldots, a_M\}$ και ένα σύνολο από K υπολογιστικούς κόμβους $C = \{c_1, c_2, \ldots, c_K\}$, ενώ το δικτυακό εύρος της σύνδεσης για κάθε υπολογιστικό κόμβο c_j στο C είναι B_j (σε bits/sec – bps) και σε κάθε κόμβο μπορούν να τοποθετηθούν περισσότερες από μια εφαρμογές με τον περιορισμό όμως ότι μια δεδομένη εφαρμογή τρέχει αποκλειστικά σε έναν υπολογιστικό κόμβο.

Εν συνεχεία ορίζουν για κάθε εφαρμογή κάποιους στόχους Ποιότητας Υπηρεσίας, οι οποίοι για την εφαρμογή a_i περιλαμβάνουν τον μέγιστο μέσο χρόνο εκτέλεσης r_i^{max} και ελάχιστη ρυθμαπόδοση x_i^{min}, ενώ η ένταση φορτίου η οποία σχετίζεται με την a_i ισούται με λ_i αιτήσεις το δευτερόλεπτο. Σε επίπεδο υπολογιστικών πόρων οι ερευνητές χαρακτηρίζουν την απαίτηση σε πόρους από την εφαρμογή a_i με μια τριάδα (p_i, d_i, n_i), με p_i να είναι ο υπολογιστικός χρόνος της εφαρμογής σε έναν από τους υπολογιστικούς χρόνους στο C, d_i η απαίτηση της εφαρμογής σε δευτερόλεπτα χρόνου υπηρεσίας του δίσκου και n_i το διαδικτυακό εύρος μετρούμενο σε bps, ενώ θεωρούν ότι όλοι οι υπολογιστικοί κόμβοι έχουν την ίδια χωρητικότητα.

Ένας από τους μελλοντικούς στόχους της έρευνάς μας θα αποτελέσει η συνδυαστική μελέτη επιπρόσθετων παραμέτρων σχετιζόμενες με τους πόρους όπως είναι η μνήμη, η επεξεργαστική ισχύς καθώς και το κόστος όσον αφορά στις ιδιαίτερες ανάγκες των βιομετρικών συστημάτων. Επιπροσθέτως, σκοπεύουμε να μην εμμείνουμε στην τοποθέτηση του Πλέγματος στη βιομετρική έρευνα μόνο ως ένα σύνολο τεχνολογιών οι οποίες προσφέρουν καλύτερη χρησιμοποίηση κάποιων υπολογιστικών πόρων, αλλά να μελετήσουμε τον τρόπο με τον οποίο μπορούν να

εκμεταλλευτούν ιδιαιτέρως ανεπτυγμένους μηχανισμούς διαχείρισης ροών εργασίας, διαχείρισης δεδομένων, ασφάλειας και εμπιστοσύνης σε περιβάλλοντα πλέγματος.

Επιπροσθέτως, η έρευνά μας θα επικεντρωθεί στην πολυπαραμετρική διαχείριση αντιγράφων με έμφαση στην Ποιότητα Υπηρεσίας. Συγκεκριμένα, θα αναλύσουμε τον όρο Ποιότητα Υπηρεσίας για το ευρύ φάσμα των βιομετρικών συστημάτων και θα μελετήσουμε την ταυτόχρονη ικανοποίηση πολλαπλών κυμαινόμενων απαιτήσεων οι οποίες αφορούν στην παρεχόμενη ποιότητα υπηρεσίας, ανάγοντας το πρόβλημα αυτό σε ένα πολυπαραμετρικό πρόβλημα βελτιστοποίησης.

Γλωσσάριο

Στον παρακάτω πίνακα παρατίθενται οι όροι που χρησιμοποιήθηκαν στη Διατριβή:

API	Application Programming Interface
BEM	Biometric Evaluation Methodology
BIR	Biometric Identification Record
BLR	Biometric Live Record
BMG	Biometrics Grid
CER	Crossover Error Rate
DNA	Deoxyribonucleic acid
EER	Equal Error Rate
EM	Expectation Maximization
EPIC	Electronic Privacy Information Centre
EPSRC	Engineering and Physical Sciences Research Council
FAR	False Acceptance Rate
FER	Failure to Enrol Rate
FERET	Facial Recognition Technology
FIDIS	Future of Identity in the Information Society
FMR	False Match Rate
FN	False Negative
FNMR	False Non-Match Rate
FRR	False Rejection Rate
FP	False Positive
GAR	Genuine Accept Rate
GMM	Gaussian Mixture Model
HMM	Hidden Markov Models
ISECOM	Institute for Security and Open Methodologies
ML	Maximum Likelihood

MRTDs	Machine Readable Travel Documents
OAE	Otoacoustic Emission
QoS	Quality of Service
ROC	Receiver Operating Characteristic
SLA	Service Level Agreement
SVM	Support Vector Machines
TN	True Negative
TP	True Positive
URL	Unified Resource Locator
VNTR	Variable Number of Tandem Repeats
XM2VTSDB	Extended Multimodal *Verification for Teleservices and Security* applications Database

Βιβλιογραφικές Αναφορές

[1] Jain, A. K.; Ross, Arun; Prabhakar, Salil (January 2004), "An introduction to biometric recognition", IEEE Transactions on Circuits and Systems for Video Technology 14th (1): 4–20.

[2] Bromme, A., A classification of biometric signatures, Multimedia and Expo, 2003, ICME '03, Proceedings. 2003 International Conference on Publication Date: 6-9 July 2003 volume: 3, page(s): III- 17-20 vol.3.

[3] The Biometric Consortium, Introduction to Biometrics, available at: http://www.biometrics.org/introduction.php, accessed on 29th April 2009.

[4] EPSRC, Otoacoustic Emission Based Biometric Systems, available at: http://gow.epsrc.ac.uk/ViewGrant.aspx?GrantRef=EP/E015522/1, accessed on 29th April 2009.

[5] Hildebrandt Mireille (2005); Privacy and Identity; in: Erik Claes and Antony Duff (ed.), Privacy and the Criminal Law, proceedings of the Conference on Privacy and the Criminal Law 14th-15th May 2004.

[6] Ricoeur Paul (1992), Oneself as another, Chicago and London: University of Chicago Press 1992.

[7] Claessens J., C. Díaz, S. Nikova, B. De Win, C. Goemans, M. Loncke, V. Naessens, S. Seys, B. De Decker, J. Dumortier, and B. Preneel (2003); "Applications Requirements for Controlled Anonymity," APES Deliverable D7, 129 pages, 2003.

[8] Durand Andre (2002); Three Tiers of Identity; Digital Identity World, March 16, 2002, available at : http://www.digitalidworld.com/print.php?sid=26

[9] Hildebrandt Mireille James Backhouse eds. (2005); "Descriptive analysis and inventory of profiling practices"; FIDIS deliverable 7.2, June 2005

[10] Janeen Renaghan, "Etched in Stone," Zoogoer, August 1997, (Smithsonian National Zoological Park, 26 January 2005).

[11] "Dermatoglyphics," Hand Analysis, International Institute of Hand Analysis, 24 January 2005.

[12] Z. McMahon, Biometrics: History, Indiana University, Indiana University Computer Science Department, 24 January 2005, available at http://www.cs.indiana.edu/~zmcmahon/biometrics-history.htm

[13] Alphonse Bertillon, access on 4th May 2009, available at: http://en.wikipedia.org/wiki/Bertillon

[14] Bonsor, K.. "How Facial Recognition Systems Work", accessed on 4th May 2009, available at http://electronics.howstuffworks.com/facial-recognition.htm

[15] Mansfield, T. et al. "Biometric Product Testing Final Report", UK Biometrics Working Group, 2001, http://www.cesg.gov.uk/technology/biometrics/media/Biometric_Test_Report_pt1.pdf

[16] von Hardenberg, I., "Warum Neugeborene mehr wissen, als Große manchmal ahnen", GEO (7), pp.27-42, Hamburg, 2001.

[17] Amberger, M., Fischer, S., Rößler, J., "Biometrische Verfahren – Studie zu State of the Art", Erlangen, 2003, available at: http://www.wi3.unierlangen.de/forschung/Biometrie_StateOfTheArt/Biometrie.pdf

[18] Mark Gasson, Martin Meints, Kevin Warwick, "D3.2: A study on PKI and biometrics", 2005, available at : http://www.fidis.net/resources/deliverables/hightechid/#c1785

[19] Webster, J.G., ed., Medical Instrumentation, 2nd ed. Boston: Houghton Mifflin Company, 1992, Chapter 5, pp. 249-250.

[20] Hirsch, W. and J.U. Schweichel, "Morphological Evidence Concerning the Problem of Skin Ridge Formation," Journal of Mental Deficiency Research, Vol. 17, pp. 58-72, 1973.

[21] Locard, E., "Les Pores et L'Identification des Criminels," Biologica, revue scientifique de medecin, Vol. 22, pp. 357-362, 1912.

[22] Dorte Hammelev, Dean Madden, Søren Nørby, Jill Turner, DNA profiling,European Initiative for Biotechnology Education (EIBE), 1998.

[23] Jobling, M. A., Gill, P., "Encoded evidence: DNA in forensic analysis", Nature (5), pp.739-52, 2004.

[24] Benecke, M., "Genetischer Fingerabdruck in Enzyklopädie Naturwissenschaft und Technik", 2. Auflage, 6. Ergänzungslieferung, ecomed Verlagsgesellschaft AG & Co. KG, Landsberg, 2001. See http://www.benecke.com/dna.pdf

[25] R. Brunelli and T. Poggio, "Face Recognition: Features versus Templates," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 15, No. 10, pp.1042-1052, 1993.

[26] Ming-Hsuan Yang, David J. Kriegman, Narendra Ahuja, "Detecting Faces in Images: A Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 1, pp. 34-58, Jan. 2002.

[27] K. M. Lam and H. Yan, "An Analytic-to-Holistic Approach for Face Recognition based on a Single Frontal View," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 20, No. 7, pp.673-686, 1998.

[28] Tolba, El-Baz, El-Harby, "Face Recognition – A Literature Review", International Journal of Signal Processing, Vol.2, No.2, 2005, pp.88-103.

[29] Murray, M., 'Gait as a total pattern of movement', American Journal of Physical Medicine, Vol. 46, No 1, 1967, pp. 290–332.

[30] Philips, P. J., Sarkar, S., Robledo, I., Grother, P., and Bowyer, K., 'Baseline results for the challenge problem of human ID using gait analysis', IEEE International Conference on AutomaticFace and Gesture Recognition, Washington, DC, USA, May 2002, pp. 130–135.

[31] Collins, R., Gross, R., and Shi, J., 'Silhouette-based human identification from body shape and gait', IEEE InternationalConference on Automatic Face and Gesture Recognition, Washington, DC, USA, May 2002, pp. 351–356.

[32] Hayfron-Acquah, J. B., Nixon, M. S., and Carter, J. N., 'Recognising human and animal movement by symmetry', Proceedings IEEE International Conference on Image Processing (ICIP '01), Vol. 3, Thessaloniki, Greece, October 2001, pp. 290–293.

[33] He, Q., Debrunner, C., 'Individual recognition from periodic activity using hidden Markov models', IEEE Workshop on HumanMotion (HUMO '00), Austin, Tex, USA, December 2000, pp. 47–52.

[34] Niyogi, S., and Adelson, E., 'Analyzing and recognizing walking figures in XYT', IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '94), Seattle,Wash, USA, June 1994, pp. 469–474.

[35] Haiping Lu , K. N. Plataniotis , A. N. Venetsanopoulos, A Layered Deformable Model for Gait Analysis, Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, p.249-256, April 10-12, 2006

[36] Haiping Lu , Konstantinos N. Plataniotis , Anastasios N. Venetsanopoulos, A full-body layered deformable model for automatic model-based gait recognition, EURASIP Journal on Advances in Signal Processing, v.8 n.1, p.1-13, January 2008

[37] Kale, A., Rajagopalan, A. N., Cuntoor, N., and Kruger, V., 'Gait based recognition of humans using continuous HMMs', IEEE International Conference on Automatic Face and Gesture Recognition, Washington, DC, USA, May 2002, pp. 336-341.

[38] Ming-Hsu Cheng, Meng-Fen Ho, Chung-Lin Huang: Gait analysis for human identification through manifold learning and HMM. Pattern Recognition 41(8): 2541-2553 (2008)

[39] Chen, C., Liang, J., Zhao, H., Hu, H., Tian, J., Factorial HMM and Parallel HMM for Gait Recognition, SMC-C(38), No. 1, January 2008, pp. 114-123.

[40] Yannopoulos, A., Andronikou, V., & Varvarigou, T. (2008). Behavioural biometric profiling and Ambient Intelligence. In M. Hildebrandt, & S. Gutwirth, Profiling the European Citizen : Cross-Disciplinary Perspectives. Springer.

[41] G. Dimauro, S. Impedovo, M. G. Lucchese, R. Modugno, G. Pirlo, "Recent Advancements in Automatic Signature Verification," iwfhr, pp.179-184, Ninth International Workshop on Frontiers in Handwriting Recognition (IWFHR'04), 2004.

[42] Santos, C., Justino, E. J. R., Bortolozzi, F., Sabourin, R., ‘An Off-Line Signature Verification Method based on the Questioned Document Expert’s Approach and a Neural Network Classifier’, 9th International IEEE Workshop on Frontiers in Handwriting Recognition, 2004, pp. 498-502.

[43] Miguel A. Ferrer, Jes?s B. Alonso, Carlos M. Travieso, "Offline Geometric Parameters for Automatic Signature Verification Using Fixed-Point Arithmetic," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 6, pp. 993-997, June 2005.

[44] Y. Qiao, J. Liu and X. Tang, "Offline Signature Verification using Online Handwriting Registration," Association for Computing Machinery, Inc. CVPR, pp. 1--8, June 2007.

[45] A. C. Ramachandra, K. Pavithra, K. Yashasvini, K. B. Raja, K. R. Venugopal, Lalit M. Patnaik: Offline signature authentication using cross-validated graph matching. Bangalore Compute Conf. 2009: 7

[46] Z. Quan and K. Liu, "Online Signature Verification based on the Hybrid HMM/ANN model," International Journal of Computer Science and Network Security, vol. 7, pp. 313--322, March 2007.

[47] Lei, H., and Govindaraju, V., ‘A Study on the Consistency of Features for On-Line Signature Verification’, Pattern Recognition Letters, Vol. 26, 2005, pp. 2483-2489.

[48] Julian Fierrez , Javier Ortega-Garcia , Daniel Ramos , Joaquin Gonzalez-Rodriguez, HMM-based on-line signature verification: Feature extraction and signature modeling, Pattern Recognition Letters, v.28 n.16, p.2325-2334, December, 2007.

[49] Lv, H., Wang, W., Wang, C., Zhuo, Q., ‘Off-line Chinese signature verification based on support vector machines’, Pattern Recognition Letters, Vol. 26, 2005, pp. 2390–2399.

[50] Justino, E. J. R., Bortolozzi, F., Sabourin, R., 'A comparison of SVM and HMM classifiers in the off-line signature verification', Pattern Recognition Letters, Vol. 26, 2005, pp. 1377–1385.

[51] Igarza, J. J., Goirizelaia, I., Espinosa, K., Hernáez, I., Méndez, R., Sánchez, J., 'Online Handwritten Signature Verification Using Hidden Markov Models', Lecture Notes in Computer Science, Springer, 2003, pp.391-399.

[52] Vu Nguyen, Michael Blumenstein, Vallipuram Muthukkumarasamy, Graham Leedham, "Off-line Signature Verification Using Enhanced Modified Direction Features in Conjunction with Neural Classifiers and Support Vector Machines," icdar, vol. 2, pp.734-738, Ninth International Conference on Document Analysis and Recognition (ICDAR 2007) Vol 2, 2007.

[53] ΕΦΗΜΕΡΙΣ ΤΗΣ ΚΥΒΕΡΝΗΣΕΩΣ ΤΗΣ ΕΛΛΗΝΙΚΗΣ ΔΗΜΟΚΡΑΤΙΑΣ, ΤΕΥΧΟΣ ΔΕΥΤΕΡΟ Αρ. Φύλλου 650, 9 Απριλίου 2009

[54] Hansen Marit and Andreas Pfitzmann (2008), Anonymity, Unobservability, Pseudonymity, and Identity Management - A Proposal for Terminology, 2008. available at http://dud.inf.tu-dresden.de/Literatur_V1.shtml

[55] ISO (1999); ISO IS 15408, 1999; http://www.commoncriteria.org/.

[56] Thierry Nabeth, Mireille Hildebrandt, D 2.1: Inventory of topics and clusters, 2005.

[57] Claessens J., C. Díaz, S. Nikova, B. De Win, C. Goemans, M. Loncke, V. Naessens, S. Seys, B. De Decker, J. Dumortier, and B. Preneel (2003); "Applications Requirements for Controlled Anonymity," APES Deliverable D7, 129 pages, 2003.

[58] Hampapur, A. (2005). Smart Video Surveillance. IEEE Signal Processing Magazine: 38-50.

[59] Foresti, G. (2005). Active Video-Based Surveillance System. IEEE Signal Processing Magazine: 25-37.

[60] Andronikou, V., Demetis, D., & Varvarigou, T. (2009). Biometric Implementations and the Implications for Security and Privacy. In M N Bhavani, Biometrics Techno-legal Issues. The Icfai University Press.

[61] BioID, About FAR, FRR and EER, available at: http://www.bioid.com/sdk/docs/About_EER.htm

[62] FERET, available at: http://www.frvt.org/FERET/default.htm

[63] XM2VTSDB , available at: http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/

[64] Daniel Solove. A taxonomy of privacy. University of Pennsylvania Law Review, 154(3):477–560, 2006.

[65] Clarke Roger, Introduction to Dataveillance and Information Privacy, and Definitions of Terms, 2006, available at: http://www.rogerclarke.com/DV/Intro.html

[66] Marek Kumpošt, Vashek Matyáš, "D13.6: Privacy modelling and identity", FIDIS NoE, 2007.

[67] "The Identity Project." from http://is.lse.ac.uk/IDcard/.

[68] Dhillon, G. and J. Backhouse (2000). "Information System Security Management in the New Millennium." Communications of the ACM 43(7).

[69] Uludag, U. and A. K. Jain (2004). Attacks on biometric systems: a case study in fingerprints. Proc. SPIE-EI 2004, San Jose, CA.

[70] How to fake fingerprints?, available at: http://www.ccc.de/biometrie/fingerabdruck_kopieren?language=en

[71] Matsumoto, T., H. Matsumoto, et al. (2002). Impact of Artificial Gummy Fingers on Fingerprint Systems. Optical Security and Counterfeit Deterrence Techniques IV, Proceedings of SPIE.

[72] M. Gasson, M. Meints and K. Warwick D3.2: A study on PKI and Biometrics. pp 1-138, 2005.

[73] Solms, R. v. (1998). "Information security management(1): why information security is so important." Information Management & Computer Security 6(4): 174-177.

[74] Knight, W. (2005). "Computer Crime boom costs UK millions." New Scientist, available at: http://www.newscientist.com/article.ns?id=dn7233.

[75] Marks, P. (2005). "Attempted cyber-heist raises keylogging fears.", available at: http://www.newscientist.com/article.ns?id=dn7168.

[76] Li, H., G. Hing, et al. (2000). BS7799: A Suitable Model for Information Security Management. AMCIS, Long Beach, California.

[77] Banisar, D. (2000). Privacy and Human Rights: An International survey on privacy laws and developments. Washington, Electronic Privacy Information Centre.

[78] EPIC, http://www.epic.org/

[79] Hyu-Bong Chung, Information Privacy: A Key Challenge of Information Society, available at: http://www.apiicc.org/apiicc/Lecture/IT_HRD/10.pdf

[80] USTreasury. (2005). "The use of technology to combat Identity Theft - Report on the study conducted pursuant to section 157 of the Fair and Accurate Credit Transactions Act of 2003.", available at:

http://www.treas.gov/offices/domestic-finance/financialinstitution/cip/biometrics_study.pdf.

[81] Cavoukian, A. (1999). "Consumer Biometric Applications: A Discussion Paper.", available at: http://www.ipc.on.ca

[82] Woodward, J. (1997). "Biometrics: Privacy's Foe or Privacy's Friend?" Proceedings of the IEEE 85(9).

[83] Arndt, C. (2005). The loss of privacy and identity. Biometric Technology Today.

[84] FIDIS, http://www.fidis.net/

[85] FIDIS, Budapest Declaration on Machine Readable Travel Documents (MRTDs), available at: http://www.fidis.net/press-events/press-releases/budapest-declaration-greek/

[86] PHR. (2004). "Threats to Privacy." from www.privacyinternational.org/threats2004

[87] Davies, S. (2002). "A Year After 9/11: Where are we now?" Communications of the ACM 45(9).

[88] Van Blarkom, G.W., Borking, J.J., Olk, J.G.E., Handbook of Privacy and Privacy-Enhancing Technologies – The case of Intelligent Software Agents, College bescherming persoonsgegevens, 2003, 33 p.

[89] Oliver, G. M., A study of the use of biometrics as it relates to personal privacy concerns, July 31, 1999, available at http://faculty.ed.umuc.edu/~meinkej/inss690/oliver/Oliver-690.htm

[90] Brown, D., Brook, D., 'Biometrics: Implications and Applications for Citizenship and Immigration - Report on a Forum hosted by Citizenship and Immigration Canada', October 7 & 8, 2003 – Ottawa, Ontario.

[91] Hildebrandt, M., Backhouse, J., Descriptive analysis and inventory of profiling practices, FIDIS Deliverable 7.2, European Union IST FIDIS Project, 2005, available at http://www.fidis.net/fileadmin/fidis/deliverables /fidis-wp7-del7.2.profiling_practices.pdf

[92] Ratha N. K., J. H. Connell, and R. M. Bolle, 'Enhancing security and privacy in biometrics-based authentication systems', IBM Systems Journal, 40(3), pp. 614–634, 2001;

[93] Ang R., R. Safavi-Naini, L. McAven, 'Cancelable Key-Based Fingerprint Templates', ACISP 2005, pp. 242-252, 2005;

[94] Cheung K. H., Ad. Wai-Kin Kong, D. Zhang, M. Kamel, Jane You, Ho-Wang Lam, 'An Analysis on Accuracy of Cancellable Biometrics Based on BioHashing', Lecture Notes in Computer Science, Springer, 2005, 1168-1172.

[95] Toh, K.-A., Lee, Ch., Choi J.-Y., and Kim J., Performance based revocable biometrics, Industrial Electronics and Applications, 2007. ICIEA 2007, 2nd IEEE Conference on 23-25 May 2007, 647 – 652;

[96] Boult, T. E., Scheirer, W. J., Woodworth, R., 'Revocable Fingerprint Biotokens: Accuracy and Security Analysis', Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on 17-22 June 2007, 1 – 8.

[97] Soutar, C., Roberge, D., Stoianov, Al., Gilroy, R., Kumar, B.V.K. V., 'Biometric Encryption', chapter 22 in ICSA Guide to Cryptography, edited by Randall K. Nichols, McGraw-Hill, 1999.

[98] A. Cavoukian, Privacy and Biometrics, Information and Privacy Commissioner, Ontario, Toronto, 1999, available at http://www.pco.org.hk/english/infocentre/files/cakoukian-paper.doc.

[99] Wikipedia, Security, accessed on 10th May 2009, available at: http://en.wikipedia.org/wiki/Security

[100] http://www.law.cornell.edu/uscode/html/uscode44/usc_sec_44_00003542----000-.html

[101] Els Kindt, Lorenz Müller, "D3.10: Biometrics in identity management", 2007

[102] Anlong Ming and Huadong Ma, "The Biometrics Grid: A Solution to Biometric Technologies," IEEE Distributed Systems Online, vol. 8, no. 9, 2007, art. no. 0709-o9001

[103] Karthik Nandakumar, Yi Chen, Sarat C. Dass, Anil K. Jain: Likelihood Ratio-Based Biometric Score Fusion. IEEE Trans. Pattern Anal. Mach. Intell. 30(2): 342-347 (2008)

[104] M. Figueiredo and A. K. Jain, "Unsupervised Learning of Finite Mixture Models," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 3, pp. 381–396, March 2002.

[105] McLachlan G.J. and Basford K.E. 1988. Mixture Models: Inference and Applications to Clustering. Marcel Dekker, New York.

[106] D. Peel and G.J. McLachlan. Robust mixture modelling using the t distribution. Statistics and Computing, 10(4):339–348, 2000.

[107] McLachlan, G.J. & Krishnan, T. (1997). The EM algorithm and extensions. New York: John Wiley and Sons.

[108] Kent J.T., Tyler D.E., and Vardi Y. 1994. A curious likelihood identity for the multivariate t-distribution. Communications in Statistics – Simulation and Computation 23: 441–453.

[109] E. L. Lehmann and J. P. Romano, Testing Statistical Hypotheses. Springer, 2005.

[110] Nat'l Inst. of Standards and Tech nology, *NIST Biometric Scores Set—Release1*,http://www.itl.nist.gov/iad/894.03biometricscores, 2004.

[111] Goel S., and Buyya, R. (2006), Data replication strategies in wide-area distributed systems," in Enterprise Service Computing: From Concept to Deployment, R. G. Qiu, Ed. Hershey, PA, USA: Idea Group Inc., 2006, pp. 211–241.

[112] Saito Y. and Shapiro M. (2005). Optimistic replication, ACM Computing Surveys, vol. 37, no. 1, pp. 42–81, March 2005.

[113] Lei, M., Vrbsky, S. V., (2006). A data replication strategy to increase data availability in data Grids, in Proceedings of the 2006 International Conference on Grid Computing and Applications, Las Vegas, Nevada, USA, 2006.

[114] Lei, M., Vrbsky, S. V. and Hong, X., (2008) An on-line replication strategy to increase availability in data Grids, Future Generation Computer Systems, vol. 24, no. 2, pp. 85–98, February 2008.

[115] Lamehamedi, H., Szymanski, B., Shentu, Z., and Deelman, E. (2002). Data replication strategies in Grid environments, in Proceedings of the 5th International Conference on Algorithms and Architecture for Parallel Processing, Bejing, China, October 2002.

[116] Cameron, D. G., Carvajal-Schiaffino, R., Millar, A. P., Nicholson, C., Stockinger, K. and Zini, F., (2003) Evaluating scheduling and replica optimisation strategies in OptorSim, in Proceedings of the 4th International Workshop on Grid Computing, 2003.

[117] Chang R.-S and Chen, P.-H. (2007) Complete and fragmented replica selection and retrieval in data Grids, Future Generation Computer Systems, vol. 23, no. 4, pp. 536–546, May 2007.

[118] Lin, H., Abawajy, J. H. and Buyya, R. (2006). Economy-based data replication broker, in Proceedings of the 2nd IEEE International Conference on e-Science and Grid Computing, Amsterdam, Netherlands, December 2006.

[119] Guy, L., Kunszt, P., Laure, E., Stockinger, H., and Stockinger, K. (2002). "Replica management in data Grids, Technical Report, GGF5, July 2002.

[120] Takizawa, S., Takamiya, Y., Nakada, H. and Matsuoka, S. (2005). A scalable multi-replication framework for data Grid, in Proceedings of the 2005 Symposium on Applications and the Internet Workshops, 2005.

[121] Torres, M., Goldman, A. and Barrera, J. (2003). A parallel algorithm for enumerating combinations, in Proceedings of the International Conference on Parallel Processing, October 2003, pp. 581–588.

[122] S. G. Akl, "Adaptive and optimal parallel algorithms for enumerating permutations and combinations," The Computer Journal, vol. 30, no. 5, pp. 433–436, October 1987.

[123] S. G. Akl, D. Gries, and I. Stojmenovic, "An optimal parallel algorithm for generating combinations," Information Processing Letters, vol. 33, no. 3, pp. 135–139, November 1989.

[124] Z. Kokosinski, "Algorithms for unranking combinations and their applications," in Proceedings of the 7th International Conference Parallel and Distributed Computing and Systems. IASTED, 1995, pp. 216–224.

[125] B. B. Zhou, R. P. Brent, X. Qu, and W. F. Liang, "A novel parallel algorithm for enumerating combinations," in Proceedings of the International Conference on Parallel Processing, vol. 2, August 1996, pp. 70–73.

[126] Qiu, L., Padmanabhan, V. N. and Voelker, G. M. (2001). On the placement of Web server replicas, in Proceedings of the IEEE Infocom 2001, vol. 3, Anchorage, AK, USA, April 2001, pp. 1587–1596.

[127] Cronin, E., Jamin, S., Jin, C., Kurc, A. R., Raz, D. and Shavitt, Y. (2002). Constrained mirror placement on the Internet, IEEE Journal on Selected Areas in Communications, vol. 20, no. 7, pp. 1369–1382, September 2002.